


```

|||||
259 CAGTAACCTTATTCGTGTCAGAACTTACTGTTATGTTAGTGA 308
101 AsphrngluthrGlnLysLysLeuProGluArgLysTrpAlaIleValTh 117
|||||
309 GATAGTGAAGTCAAAAAAGTTACTGAAAGAAATGGCGTATCGTTAC 358
117 rSerGlySerProTyrLeuAlaPheSerTrpPheGluThrIleLeuLysA 134
|||||
359 CTCGTGTTCCCATATTTGGATTTCATGTTGACGACATATTTGAAAA 408
134 snValGlyLysProLysValPheIleThrGlyPheAspValLysAsnGly 150
|||||
409 ATGTGGAAAGCCCAAGTTTTCATCTACTGGGTTTGACGTGAAGACGGT 458
151 LysProAspProGluGlyTyrSerArgAlaArgAspLeuLeuArgGlnAs 167
|||||
459 AAGCCTGATCCCGAGGGTTATTCAAGAGCTCGTATTATTGCGTCAAGA 508
167 PleuGlnLeuThrGlyLysGlnAspLeuLysTyrValValPheGluAspA 184
|||||
509 TTTCGATTAATCTGTATGACAGAGATCTGAAGTATGTTGCTTCAGAGATG 558
184 IarProValGlyIleLysAlaGlyLysAlaMetGlyAlaIleThrValGly 200
|||||
559 CACCCGGGGCATAAAGCGCGCAAAACCAATGGCGCATTAAGTGGGT 608
201 IleThrSerSerTyrAspLysSerValLeuPheAspAlaGlyAlaAspTy 217
|||||
609 ATAAACATCCCTCGTATGACAGAGCGCTTTTATTGACCGACAGACAGTTA 658
217 rValValLysAspLeuThrGlnValSerValValLysAsnAsnGluAsnG 234
659 TGTAGTCTGTGATTGACACAGGTTCCGTGTTAAGAACATGAACAAACG 708
234 LylIleValIleGlnValAsnAsnProLeuThrArgAla 246
709 GTATTGTGATCCAGGTAAACACCTTTTGACAAAGGCC 746

```

seq_name: gb_pat1:A85030

seq_documentation_block:

LOCUS A85030 758 bp DNA

DEFINITION Sequence 1 from Patent EP0870836.

ACCESSION A85030

VERSION A85030.1 GI:6733776

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

SOURCE

CDS

BASE COUNT

ORIGIN

```

1. 758
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
9. 749
/note="unnamed protein product"
/codon_start=1
/db_xref="GI:6733777"
/translation="MAFESADLCFLDGTIVSTYVAAEAMTKLCYEVDPSELFK
HSRGARQEVLRFRFKLDPTDNKGVALAEKDIASHSLDVTSLIPGAENLLSLDVP
ETGKKLPERKMAIVTSGSPYLASFWEETILKNVGPVFLTGFVDVKNKGPDPESYSA
RDLLRODLQITGADLVKVFEDAPVGIKAGKAMGATTVGITSYDYSVLFDAAGADV
VCDLTGVSVYKNNNGIVIOVNPDLTPA"
221 a 135 c 183 g 219 t

```

alignment_scores:
 quality: 1269.00 length: 246
 Ratio: 5.159
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-558-284-2 x A85030 ..
 Align seg 1/1 to: A85030 from: 1 to: 758

```

1 MetAlaGluPheSerAlaAspLeuGlyLeuPheAspLeuAspGlyThrI 17
|||||
9 ATGGCAGAAATTTTCACTGATCTATCTTTTGGACCTAGATGTCACAT 58
17 eValSerThrThrValAlaAlaGluLysAlaTrpPheLysLeuGlyTyrG 34
59 AGTAGAGACAAACAGTGGCCGACAGAAAGCATGACCAAGTTGCTTACG 108
34 IutYrGlyValAspProSerGluLeuPheLysHisSerHisGlyAlaArg 50
109 AATACGGTGTTCATCCTCCGAGTTATTAMGCATTTCTCATGGTGAAGA 158
51 ThrGlnGluValLeuArgArgPhePheProLysLeuAspAspThrAspAs 67
159 ACACAGAGAGGTTTGAAGAGGTTTTCCTAAATTGGATGATACAGACAA 208
67 nLysGlyValLeuAlaLeuGluLysAspIleAlaHisSerTyrLeuAspT 84
209 TAAAGGTGTTCTTCTCTAGAAAAAGATATTGCCCATAGTTACTTGATA 258
84 hrValSerLeuIleProGlyAlaGluAsnLeuLeuSerLeuAspVal 100
259 CAGTAAAGCTTATTCTCTGGTGACAGAGACTTACTGTTATCGTTAGATGA 308
101 AspThrGluThrGlnLysLysLeuProGluArgLysTrpAlaIleValTh 117
309 GATAGTGAAGTCAAAAAAGTTACTGAAAGAAATGGCGTATCGTTAC 358
117 rSerGlySerProTyrLeuAlaPheSerTrpPheGluThrIleLeuLysA 134
359 CTCGTGTTCCCATATTTGGATTTCATGTTGACGACGACAAATTTGAAAA 408
134 snValGlyLysProLysValPheIleThrGlyPheAspValLysAsnGly 150
409 ATGTGGAAAGCCCAAGTTTTCATTAAGGTTTGACGTGAAGACGGT 458
151 LysProAspProGluGlyTyrSerArgAlaArgAspLeuLeuArgGlnAs 167
459 AAGCCTGATCCCGAGGGTTATTCAAGAGCTCGTATTATTGCGTCAAGA 508
167 PleuGlnLeuThrGlyLysGlnAspLeuLysTyrValValPheGluAspA 184
509 TTTCGATTAATCTGTAAACAGAGATCTGAAGTATGTTGCTTGAAGATG 558
184 IarProValGlyIleLysAlaGlyLysAlaMetGlyAlaIleThrValGly 200
559 CACCCGGGGCATAAAGCGCGCAAAACCAATGGCGCATTAAGTGGGT 608
201 IleThrSerSerTyrAspLysSerValLeuPheAspAlaGlyAlaAspTy 217
609 ATAAACATCCCTCGTATGACAGAGCGCTTTTATTGACGACGACGAGATTA 658
217 rValValLysAspLeuThrGlnValSerValValLysAsnAsnGluAsnG 234
659 TGTAGTCTGTGATTGACACAGGTTCCGTGTTAAGAACATGAACAAACG 708
234 LylIleValIleGlnValAsnAsnProLeuThrArgAla 246
709 GTATTGTGATCCAGGTAAACACCTTTTGACAAAGGCC 746

```

seq_name: gb_p13:SCU03107

```

seq_documentation_block:
LOCUS      SCU03107      1650 bp      DNA      15-OCT-1994
DEFINITION Saccharomyces cerevisiae 2-deoxyglucose-6-phosphate phosphatase
              (DGC1) gene, complete cds.
ACCESSION  U03107
VERSION    U03107.1  GI:495309
KEYWORDS
SOURCE     Saccharomyces cerevisiae.
            Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE  1 (bases 1 to 1650)
            Sanz,P., Rander-Gill,F. and Prieto,J.A.
            Molecular characterization of a gene that confers 2-deoxyglucose
            resistance in yeast
            Yeast 10 (9), 1195-1202 (1994)
JOURNAL    95274321
MEDLINE
REFERENCE  2 (bases 1 to 1650)
            Sanz,P.
            Direct Submission
            Submitted (03-NOV-1993) Pascual Sanz, Instituto Agroquimica y
            Tecnologia Alimentos (CSIC), Cereales, Jaime Roig, 11, Valencia,
            Valencia, 46010 Spain
            Location/Qualifiers
            1..1650
               /organism="Saccharomyces cerevisiae"
               /db_xref="taxon:4932"
               /clone="10a12-13"
            649..1389
               /gene="DGC1"
               /locus_tag="Y03107"
               /protein_id="AAC48923.1"
               /db_xref="GI:495310"
               /translation="MAEFSADGLFDLDQITVTAAEKAMTKLCYEGVDSPELFK
               HSHGARQGEVLRFPKLDPTDNKGVLALEKDIASHYLDVSLIQAEMLALSLQVDI
               ETOKLPERKVAIVMSGSRYLAFSMEFTILKAVGPKRVTITGPDVKNKNGPDEGTSRA
               RDLRLQDILQITGKODLKYVFEDAVGIIKAGAMGATIVGITSIDKSVLPDAGADYV
               VCDLTVSVVKNNGIVIVONNPLTRA"
BASE COUNT 501 a      280 c      333 g      536 t
ORIGIN
alignment_scores:
      Quality: 1269.00      Length: 246
      Ratio: 5.159      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-558-284-2 x SCU03107 ..
Align seg 1/1 to: SCU03107 from: 1 to: 1650
1 MetAlaGluPheSerAlaAspLeuCysLeuPheAspLeuAspGlyThrIi 17
  ||||||||||||||||||||||||||||||||||||||||||||||||
649 ATGGCAGAAATTTTCAGCTGATCTATGCTTTTGGACCTACATGGTACCAT 698
17 eValSerThrThrValAlaIaGluLysAlaTyrPThrLysLeuCysTyrG 34
  ||||||||||||||||||||||||||||||||||||||||||||||||
699 AGTGCAGACAAACAGTCGCCAGAGAAAGCATGACCAAGTTCGTTACG 748
34 LuTyrgLyValAspProSerGluLeuPheLysHisSerHisGlyAlaArg 50
  ||||||||||||||||||||||||||||||||||||||||||||||||
749 AATACGCTGTGATCCTTCGAGATTATTAAGCATTCCTCATGTCGCAAGA 798
51 ThrGluGluValLeuArgArgPhePheProLysLeuAspSptThrAspAs 67
  ||||||||||||||||||||||||||||||||||||||||||||||||
799 ACACAGAGAGGTTTGGACAGAGGTTTTCCTAATTTGATGATACAGACAA 848

```

```

67 nLysGlyValLeuAlaLeuGluLysAspIleAlaHisSerTyrLeuAspT 84
  ||||||||||||||||||||||||||||||||||||||||||||||||
849 TAAAGGCTCTCTGCTCTAGAAAAAGATATTGCCCATGATTGACGACA 898
84 hrValSerLeuIleProGlyAlaGluAsnLeuLeuLeuSerLeuAspAl 100
  ||||||||||||||||||||||||||||||||||||||||||||||||
899 CAGTAAAGCCTTATTCCTGTCGACAGAACTTACTGTTATCTGTATGATGA 948
101 AspThrGluThrGluLysLysLeuProGluArgLysTyrPalIleValTh 117
  ||||||||||||||||||||||||||||||||||||||||||||||||
949 GATACGAGACTCAAAAAAGTTACTCAAGAGAAATGGCTATGCTTAC 998
117 rSerGlySerProTyrLeuAlaPheSerTyrPheGluThrIleLeuLysA 134
  ||||||||||||||||||||||||||||||||||||||||||||||||
999 CTCGTGTTCTCCCTATTGTCGATTTTCATGTCGTCAGACATATTGAAA 1048
134 snValGlyLysProLysValPheIleThrGlyPheAspValLysAsnGly 150
  ||||||||||||||||||||||||||||||||||||||||||||||||
1049 ATGTTGGAAGCCCAAGTTTTCATTTACTGGCTTTGACGTGAAGAACGCT 1098
151 LysProAspProGluGlyTyrSerArgAlaArgAspLeuLeuArgGlnAs 167
  ||||||||||||||||||||||||||||||||||||||||||||||||
1099 AAGCCTGATCCCGAGGGTTATTCAAGAGCTGCTGATTTATTGCGTCAAGA 1148
167 pLeuGluLeuThrGlyLysGlnAspLeuLysTyrValValPheGluAspA 184
  ||||||||||||||||||||||||||||||||||||||||||||||||
1149 TTTGCATTAATTAAGTGTAAACAGAGATCTGAATGTTGTTGTAAGATG 1198
184 lAprOvalGlyIleLysAlaGlyLysAlaMetGlyAlaIleThrValGly 200
  ||||||||||||||||||||||||||||||||||||||||||||||||
1199 CACCCGTGGCATTAAGCCCGCTTAAGCAATGGCGCAATTACTGTGGCT 1248
201 lIleThrSerTyrAspLysSerValLeuPheAspAlaGlyAlaAspTyr 217
  ||||||||||||||||||||||||||||||||||||||||||||||||
1249 ATACATCTCTGATGATTAAGAGCGTTTATTATGCGCAGAGCGAGATTA 1298
217 rValValLysAspLeuThrGlnValSerValValLysAsnAsnGluAsnG 234
  ||||||||||||||||||||||||||||||||||||||||||||||||
1299 TGTAGTCTGATTTGACACAGGTTTCCGCTTAAGCAATGTAAGAACG 1348
234 lYlleValIleGlnValAsnAsnProLeuThrArgAla 246
  ||||||||||||||||||||||||||||||||||||||||||||||||
1349 GTATTGTCATCCAGTAAACACCTTTGACAGAGGCC 1386

seq_name: gb_p14:YSC8179

seq_documentation_block:
LOCUS      YSC8179      4413 bp      DNA      04-SEP-1997
DEFINITION Saccharomyces cerevisiae chromosome VIII cosmid 8179.
ACCESSION  U00062 U00093
VERSION    U00062.1  GI:488162
KEYWORDS
SOURCE     Saccharomyces cerevisiae strain=5288C (AB972).
            Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE  1 (bases 1 to 4413)
            Johnston,M., Andrews,S., Brinkman,R., Cooper,J., Ding,H., Doyer,J.,
            Du,Z., Favello,A., Fulton,L., Gatlung,S., Geisel,C., Kirsten,J.,
            Kudaba,T., Hillier,L., Jier,M., Johnston,L., Keppeler,D.,
            Langston,Y., Latreille,P., Louis,E., Macri,C., Mardis,E.,
            Mouser,L., Nhan,M., Rifken,L., Riles,L., St.Peter,H., Thornton,L.,
            Trevasakis,E., Vaudin,M., Vaughan,K., Vignati,D., Wilcox,L.,
            Willis,A., Wilson,R., Woldman,P. and Waterston,R.
            Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
            VIII
            Science 265 (5181), 2077-2082 (1994)
JOURNAL    94378003
MEDLINE
PUBMED    8091229
REFERENCE  2 (bases 1 to 4413)
            Du,Z.
            The sequence of S. cerevisiae cosmid 8179
            TITLE

```

JOURNAL Unpublished (1994)
REFERENCE 3 (bases 1 to 44113)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (19-MAY-1994)
REFERENCE 4 (bases 1 to 44113)
AUTHORS Jia, Y. and Cherry, J. M.
TITLE Direct Submission
JOURNAL Submitted (04-SEP-1997) Department of Genetics, Stanford University, Saccharomyces Genome Database, Stanford, CA 94305-5120, USA

COMMENT
Curated by:
Saccharomyces Genome Database
URL: <http://genome-www.stanford.edu/>
e-mail: yeast-curator@genome.stanford.edu

Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: mj@sequencer.wustl.edu

NEIGHBORING COSMID INFORMATION:

This sequence includes nucleotides 1-43838 of cosmid 8179 plus 75 bp of cosmid 9181 (cosmid adjacent to the right end of 8179). This sequence overlaps with cosmid YSC88025 on the right by 200 bp. The adjacent cosmid to the left is YSC88082.

location/Qualifiers
1. .44113
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/map="VII1"
/strain="S288C (AB972)"
/complement(539. .1993)
/gene="SLT2"
complement(539. .1993)
/gene="SLT2"
/note="YHR030C: Other names MKP1, SLK2, BYC2."
/codon_start=1
/product="Slp2p: Putative Ser/Thr protein kinases"
/protein_id="AAB68912.1"
/db_xref="GI:488177"
/translation="MAKDIERHFKVFNODPSVYKRPQLIKETIGAGYCYCSAPAE
AAETTYAIKVTNVESEKTLCKRSLEKLIRFRGKNITTCYDMDIYYPGGSIN
GLYIEELMECDHQLIKSGPLTDAHYQSTFYQITGLKATISADVLHRLKQGNL
VNACQQLKIDCFGLARGYSENPENSOQLTEYVATRWRAPEIMLSYGATKADVMS
AGCILAFLGKRPFGKDYVNOINQLVGTPTDETLRIGSKNODYTHQLGPIF
KVPVNLTPNANSQALDLBOMLAFOKRTTVELEHPYLSIMHPADBPVCSSEK
EFSPESYNDMEDLKQWYIOYODRFLYRQPLFEORQLOLQOOOQOOOQOOOQOO
SDVNGMAAASEENYPKQMATSNVAPQOESFGIHSQNLPRHDAFPFRPQESMEER
PATGNTADIPPNQNGTLDLLEKLEFGLDKRP"

gene
complement(2448. .4619)
/gene="YHR031C"
complement(2448. .4619)
/gene="YHR031C"
/note="Similar to Pif1p (S. cerevisiae), mitochondrial DNA repair and recombination protein."
/codon_start=1
/evidence="not_experimental"
/product="Yhr031cp"
/protein_id="AAB68913.1"
/db_xref="GI:488178"
/translation="MFRSHASGNKKQWSSKSSNGSTPAASAGSHAYRQOITLSSFFMG
/CKSAASAKNSTTIIIDESGDEGNRMTAPPRRLIRNNSSSLFESQSGFGDDDD
AEFKKIVADRLNYSKSSSLSWTSLSLHTASPTQKTYHDEDETLEAVTSVKS
NSRLSTSTINIEDSSMKLSDESRPAKRSKSPMEFGIKLTPKTKIKPLIKVSNM
DSNMHRASSPVVLTMEQERYVNLIVKKRTNPEYTGSGAGCKSVIOTIIRQLSLYG
KESATATAGLAAYTIGGSTLHKWSGIGKNTIDQVKKIQSQKDLAWKRTKYL
ITDEISWQNLQKLEQIARIRKNDPFGQIQDLVTGDFQLPPPAKQDEHNVAE
CFSEMWKRCIOKTIILTKVFRODNKILDLNAIRVGEITVDIAKTIIRNLNDIDYA

gene
complement(4993. .6738)
/gene="YHR032W"
4993. .6738
/gene="YHR032W"
/codon_start=1
/evidence="not_experimental"
/product="Yhr032wp"
/protein_id="AAB68911.1"
/db_xref="GI:488176"
/translation="MSKQSHNTNDRBSIIYSTSVGAGLFTPADYIPQSENLIRE
GEEBGSSEEPSTYGNDETEREGEYHSLDANNSRTIQQAMQOQYSHRKRILDE
ERDLITDKKLISOHONGGDIESHQGAIGDEERPAELIANTWESATIESQKISTP
FKRETVYTNNALPLIFPIILQNSLSIASISVSHLGKELGAVTIGSMANTIGLAA
IQGLCTCLDLCQAQAYGAKNYHLGVLQVQCAVITILAFPMYVMEVMSKILALMI
PERELCALANAYLRTAFGVPGEILFEGCKRFLOCOGIFHASTIYLVFCAPLNMNY
LLVMNDKIGTIGLAPLSVYINVMYMLGLIYAMTTHKREPKCWNGIIPKQDAFK
NMKMINAIPIGVVVAEAEFLGPEVLTIFASHLGTDAQASIYATISLAVQYQFST
SVSTIRVANTIGASLYDSQMITCRVSLLSFVCSMMMFVICTKEDIASLFTESA
VKKWVVDTLPLAFMQLEFADFNASTAGLQGGQKQKRWVHQPSKILLPRCAHGICVS
IPVSSGCRRLMWGYN"
7197. .8468
/gene="YHR033W"
7197. .8468
/gene="YHR033W"
/note="Highly similar to Pro1p (Gamma-glutamyl kinase)"
/codon_start=1
/evidence="not_experimental"
/product="Yhr033wp"
/protein_id="AAB68910.1"
/db_xref="GI:488175"
/translation="MTKAVTVIKRGSSSLVDESTKEPKLSIMTLIVETYNLRMGH
KVIIVSSGGLAVGLDALNI PHKPKQISEVQAIAVGGQRLIARNMMLFSQGEQTAQI
LITRNDILRMQYNNARNITINELAMGYIPIVENDTSLISEIEGDDTLATATA
VGADPELITGVDCLTNDTPNDPARDPIVAVPELSEGLPGVNTSSGSGSVGGAMR
TKLIAADLASNAGETIYMKSDREYVYKIVDYIOHHRKPRRHGNGCQOQFDLODT
ELEQDRTDVPMTKFLANDKHKILKNEPILHGLITKGAIIDENSTIDLSKDMA
SLTPNAVILEVDNHFLECDLKGIRLNGEILDISKIPQSGVARSYTSLELAKIK
GLPSEKIHVDYGVSEVVAHRENIAPPOQ"
complement(8614. .9648)
/gene="YHR034C"
complement(8614. .9648)
/gene="YHR034C"
/codon_start=1
/evidence="not_experimental"
/product="Yhr034cp"
/protein_id="AAB68914.1"
/db_xref="GI:488179"
/translation="MADFLRPQIRHRENEKYVSDAAGSVSKIEPIADFYIKTKL
LSANGPERLQGRKVFVNVCHSPVPRKEDVFNARYEPILLIQNEMELPITTSYRMD
HDKKGOECYVWDCINSDCSRWICDDQILREILVEMCESECIKSDSVLCNDRYAFK
MKKGAELPALEVINDELHODYKAKMHIIEEGADPSIILGRNDDODDNDPDGFT
LPLPLENTKISGAKIEIDIKNTAHRKLQAPAPAPAPHQOQDVEPEYKMKRFG
AAVYKRLILEKAKNSKDRSPSYNRAENLIYNGKLSITLPDPIYVNNADIKFIHI
RKERTLYIT"
9868. .11760
/gene="YHR035W"
9868. .11760
/gene="YHR035W"
/note="Weak similarity to Sec23p (S. cerevisiae)."
/codon_start=1
/evidence="not_experimental"
/product="Yhr035wp"
/protein_id="AAB68909.1"
/db_xref="GI:488174"
/translation="MCSPTNPLYPFSSDAYTQWYDNQDKTKCKGAYYSNACSLREON
VMTLFCNQSNAELPLVPSNTYTLTSAKEILSKRTIMTIDICDHEINLYVSLI
KNYNTIRQOEPLSLITIQSGCHVILNANVNRRAVPSINFPMTKYNLDKINASYFEK
KISEINQESYWFDSIGSLRKLRLREPLKIANKNVNISSKRDKCTGLALFVSSVLAQ

CSLSAVCHIVFLNGPCTKGGKVMRSERGESMRONHFEESKSQOLSKSPTKFYK
MLEKFNANSLIEFFIASLDQIGILEMSPLITSSMAVSQDFSNDRFAMSPQYINL
RDHNAIVNCHSKIMTAKNAIVKDEPKYSLNKSNLSPLEISLGHNSRFAPOPTF
ENOTKRYRIERILLPKANRSFGAONEIVFSKAKTASRIIDSFAVSSKTKELMOFL
LLPNORCKRDVNMVNLICMCHYIYRSPILSVNTSPDEKYLELHRIANSKDTGLSLC
KPFISYSDLKHDWIVLDVPIITRAQI1LDDDKTICVDGGSYLVLRGKLLKEGRFLC
CKLANDIQRFPOLYVETKGTGSGSDRFLKSI1PLDITDKETLGTEDMTFNEYFLNLT
DLGSGK"
11994..13409
/gene="YHR036w"
11994..13409
/gene="YHR036w"
/codon_start=1
/evidence="not_experimental"
/product="Thr036wP"
/protein_id="AAB68908.1"
/db_xref="GI:488173"
/translation="MESTFENLSIRDSFTSGMEHVEDELGLSDLSISKQGPPLSPQLI
NRFMPHFPSSPLRNTLDFSAKADEEDDMEDIDEVDITSEEEYNEPIETHTEA
TENAVEIEIATPEEROKOEKNSODQSEVEENIVSPRSTVIALSPDLGYAAA
TKVEGVPLPSPANODNESSNNNAEGEDIINNEVEDEIKSLGHNKSDSOYANAFKL
EIKRELRSKSYOPIOVSFNTHNYEYSPDKDITKYSILTKPWHNKIDEPYDQNEAFKL
PKWSPNHPASRASAYLMSYLOLPLNATTVITVTSILSTIALOKDKLWBEORKH
ELQYBSRICQEOYLTRNCNQTTPGLPALGEOCAIMKQCMRNDIDFRASSTLSAKLFG
DIINSFIDPLNKKTLFVIFCGVITWCFSSNFLGFRVAKSYGNGIKTYPLPSPKSP
TSEETHSSMTASGDSHLLKQ"
13626..15353
/gene="PUT2"
13626..15353
/gene="PUT2"
13626..15353
CDS

alignment_scores:
Quality: 1269.00 Length: 245
Ratio: 5.159 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-558-284-2 x YSCH8179/rev ..

Align seg 1/1 to reverse of: YSCH8179 from: 1 to: 4413

1 MetAlaGluNheSerAlaAspLeuGlyLeuPheAspLeuGlyThr11 17
|||||
26457 ATGCGAAGATTTCAGCTGATGCTATTGCTTTGGACCTAGATGGTACCAT 26408
17 eValSerThrThrValAlaIaGluLysAlaTrpThrLysLeuGlyTyrG 34
|||||
26407 AGTGAAGACACAGTGGCCGACAGAAAGCATGACCAAGTTGCTTACG 26358
34 LuTrpGlyValAspProSerGluLeuPheLysHisSerHisGlyAlaArg 50
|||||
26357 AATACGCTGTGATCCTCCGACGTTATTAAACATCTCTCATGTGCAAGA 26308
51 ThrGlnGluValLeuArgArgPhePheProLysLeuAspSerThrAspAs 67
|||||
26307 ACACAGAGGCTTTTGAGAAAGCTTTTCCCTAAATTCGATGATGACAGCAA 26258
67 nLysGlyValLeuAlaLeuGluLysAspIleAlaHisSerTyrLeuAspT 84
|||||
26257 TAAAGGTGCTTCTGCTCTAGAAAAAGATATTGCCCATAGTACTTGACA 26208
84 hrValSerLeuIleProGlyAlaGluAsnLeuLeuSerLeuAspVal 100
|||||
26207 CATTAGACCTTATTCTGCTGACAGAACTTACTGTTATCTAGATGTA 26158
101 AspThrGlnThrGlnLysLysLeuProGluArgLysTrpAlaIleValTh 117
|||||
26157 GATACGACGACTCAAAAAAGTACCGAAAGCAAAAGCGCTATCGTTAC 26108
117 rSerGlySerProTyrLeuAlaPheSerTrpPheGluThrIleLeuLysA 134
|||||
26107 CTCTGGTCTCCATATTGTCATTTGTCATGTCGACACATATTGAAAA 26058
134 snValGlyLysProLysValPheIleThrGlyPheAspValLysAsnGly 150

|||||
26057 ATGTGGAAAGCCCAAGTTTCACTTACTGGGTTGACGTGAAGAAGCGT 26008
151 LysProAspProGluGluGlyTyrSerArgAlaArgAspLeuLeuArgGlnAs 167
|||||
26007 AAGCGTATCCCGAGGCTTATTCAAGAGCTCGTGAATTTATTCGCGCAAGA 25958
167 pLeuGlnLeuThrGlyLysGlnAspLeuLysTyrValValPheGluAspA 184
|||||
25957 TTTCGAATTAACGTGTAAACAGAGATCTGAATATGTTCTTCCAGAGATG 25908
184 lAprOvalGlyIleLysAlaGlyLysAlaMetGlyAlaIleThrValGly 200
|||||
25907 CACCGTGGCATTAAGCGCGCAAGCAATGGCGCATTAAGTGTGGCT 25858
201 lLehTrSerSerTyrAspLysSerValLeuPheAspAlaGlyAlaAspTy 217
|||||
25857 AATACATCTCTGATGACAAAGCGTATTATTTGACGACGAGACAGATTA 25808
217 rValValCysAspLeuThrGlnValSerValValLysAsnGluAsnG 234
|||||
25807 TGTAGTGTGATTTGACACAGGTTTCGCTGTTAAGACATGAAAAAG 25758
234 lYlIleValIleGlnValAlaAsnProLeuThrArgAla 246
|||||
25757 GTATTGTCATCCAGGTAAACAAACCTTTGACACAGGCGC 25720

seq_name: gb_sts1:CNS0616P

seq_documentation_block:

LOCUS CNS0616P 826 bp DNA STS 10-JAN-2001
DEFINITION 73 end of clone AS0A018C02 of library AS0A from strain CLIB 533
ACCESSION AL399911
VERSION AL399911.1 GI:12155045
KEYWORDS STS.
SOURCE Saccharomyces bayanus.
ORGANISM Saccharomyces bayanus

REFERENCE 1 (bases 1 to 826)
Boulet,J.L., Aigle,M., Casaregola,S., Artiguenave,F., Wincker,P.,
Bolotin-Fukuhara,M., Bon,E., Brotier,P., Casaregola,S.,
Aigle,M. and Durans,P.
Genomic Exploration of the Hemiascomycetous Yeasts: 5.
Saccharomyces bayanus var. uvarum
FEBS Lett. 487 (1), 37-41 (2000)
11152880
2 (bases 1 to 826)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brotier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durans,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
yeast species for molecular evolution studies(1)
FEBS Lett. 487 (1), 3-12 (2000)
11152876
3 (bases 1 to 826)
Genoscope.

Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This STS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of

COMMENT


```
FEATURES             source          misc_feature
misc_feature         gene           CDS
source               location/Qualifiers
    .37968
    /organism="Candida albicans"
    /strain="1161"
    /db_xref="taxon:5476"
    /clone="cosmid Ca20c1"
    96..102
    /note="possible tata box, tataaaa upstream of Ca20c1.01"
    172..894
    /gene="Dog2"
    /note="Ca20c1.01"
    172..894
    /gene="Dog2"
    /note="Ca20c1.01, Dog2, len: 240 aa, similar to a family
of proteins that include S. cerevisiae
2-deoxyglucose-6-phosphate phosphatase 2 (EC 3.1.3.-)
(2-DOG-6-P 2), DOG2 OR YHR043C, DOG2_YEAST P38773, fasta
scores opt: 476, (43.0% identity in 230 aa overlap) and
DOG1_YEAST P38774, 2-deoxyglucose-6-phosphate phosphatase
1 (EC 3.1.3.-) (2-DOG-6-P 1). DOG1 OR YHR044C (246 aa)
opt: 462, (41.4% identity in 239 aa overlap)"
    /codon_start=1
    /transl_table=12
    /product="putative 2-deoxyglucose-6-phosphate phosphatase
/protein_id="CA21923.1"
/db_xref="GI:3850122"
```

```

gene
1136. .2860
/ gene="Ca20c1.02"
1136. 2860
/ note="Ca20c1.02"
/ note="Ca20c1.02, len: 574 aa, similar to RNA helicases eg
S. cerevisiae Dbp9_YEAST Dbp9, (594 aa), fasta scores,
opt: 1568, E(): 0, (62.6% identity in 588 aa overlap);
contains P500017 ATP /GTP-binding site motif A, Pfam
match to entry DEAD PF00270, DEAD and DEAD box helicases, Pfam
match to entry helicase_C PF00271, Helicases conserved
C-terminal domain"
/codon_start=1
/transl_table=12
/product="dead box helicase"
/protein_id="CAA21924.1"
/db_xref="GI:3850123"
/translation="MSTASSSYLDDDETTPWSEFNIDPRLLOAIQDLSFNPLIQSSA
IPLEAREKDIITAKSTGSGKTAAYCIPITVNNLITDSSQGIKSTIIIVPTSLNQHVA
QVEVLILFFSTKRIKIVNLSSSYSDQVNLNLVNNKPEITISTPAKLIILKRNKNEI
LSTVNLILIDEVDILVLSGIDDLKLEESYLPVKKNIQATFLMSATVNDLDDILQKRCY
TPALIKLINDSANCNNLVQYAAKTTEDKFLAVIKRLNIKKRTAFANINDRGFLALV
RIKLELDFQGIKCCILINSEIPLNSRLHIVVEEKNKNVYLLIATDETNELSEONDENE
GDEIDKKDGNATPKRKRSKKSKKREKQDREYGSRCVDPRNVAQVINDPLPSYSKIIH
IGRTARAGAGALSFVLPISFEGHKHKTASLASAKDKDVLGRIVYQDQSKNGFELKRP
QEDNMNQVGFRRADARAVQTAVTARARAEKLNELINSEKIKRFEENPDOLASTS
RHDKLHARITQSLKNPQVYLLIPESARQDVNLTGIVFPHNKKIHKHKGKSGKSKKY
```

```
misc_feature      1193..1903
                  /gene="Ca2001.02"
                  /note="Pfam match to entry DEAD PF00270, DEAD and DEAD box
                  heliases, score 220.48"
```

```
misc_feature      1310..1333
                  /gene="Ca2OC1.02"
                  /note="PS00017 ATP /GTP-binding site motif A"
misc_feature      1961..2335
                  /gene="Ca2OC1.02"
                  /note="Pfam match to entry Helicase_C PF00271, Helicases
                  conserved C-terminal domain, score 44.51"
                  complement(2991..3620)
gene              /gene="Ca2OC1.03c"
```


CDS
complement(2991..3620)
/gene="Ca20C1.03c"
/note="Ca20C1.03c, unknown len: 228 aa, slight similarity to 014142, hypothetical 23.4 kDa protein, (206 aa), fasta scores, opt: 207, E(1): 2.6e-06, (26.6% identity in 177 aa overlap)"
/codon_start=1
/transl_table=12
/product="hypothetical protein"
/protein_id="CAA21925.1"
/db_xref="GI:3850124"
/translation="MSLPFAEKVKDADLEKSMADSKNYHLHPTSLTYKKLLFALLI
ESTISCFYTYIDRGLIHPLAPAMLCESGALASQSMNOYSRKFNKIFRWVG
FINGFSLVAMISILVTRVDNLFYRILDDSCGTMPQLFNILNALWDGGEVSSATR
IAFPLSLVASYSCFMFPFISVSPFIFRITLMPCCNLANLFMNLILSKIA"
4545..6077
/gene="Ca20C1.04"
4545..6077
/note="Ca20C1.04", unknown membrane protein, len: 510 aa,
similar to eg YHK4_YEAST P38776 hypothetical 57.8 kDa
protein in APL-SMF2 intergenic region YHR048W. (514 aa),
fasta scores, opt: 1139, (47.4% identity in 468 aa
overlap)"
/codon_start=1
/transl_table=12
/product="hypothetical membrane protein"
/protein_id="CAA21926.1"
/db_xref="GI:3850125"
/translation="MEIEFEKKKKDILLDDFTTNRVLEPELLFKTSDFYLVFDD
DDKEANSLSRHKKWTVIGLISGVCTVSSSSVALATPQIMHDIHSEVALNLS
FFILLDGGCGFLPISFEFGRIYVIGGLTMTAFELFANISICAMVGRMSGF
CGSSVMSVASSPFDLFEARKKEGDKANKELSLALLISVAPFLGPSGLISGFI
NTHLVYMFETVFLIMSGVLLAVLVLEPETHPEVNLORAKRIKRTGDNRYVAAADQ
TTSLSYESTLTSKRPIILLKDMNFALLNYGTFLAVVYLFPAFPISTSYHPTI
SEQSSFLGLVAGWSTCLISPTFDGCVALLRKNNGVAKPEPRFIPLVGAFAPVPI
GLFPIIATVSYSHIHWIAPLIGSAVYSGTFLVFNISPAVYEAIVLAASAMANSLT
RVSYLSCIFPLGFLQMYKGLGIQMATTLTLLAFACLMIPFLFFKYGGLRSRSPYAWS
VD"

misc_feature
6255..6261
/note="potential tata box, tatlaaac, upstream of
Ca20C1.05"
join(6298..6323,6407..6836)
/gene="Ca20C1.05"
join(6298..6323,6407..6836)
/note="Ca20C1.05, unknown, spliced, len: 151 aa, similar
to S. cerevisiae YN21_YEAST P53834, hypothetical 17.2 kd
protein in MRPL10-ERG24 intergenic region, YNL281W or
M0589, fasta scores, YN21_YEAST P53834 opt: 384, (39.2%
identity in 153 aa overlap)"
/codon_start=1
/transl_table=12
/product="conserved hypothetical protein"
/protein_id="CAA21927.1"
/db_xref="GI:3850126"

misc_feature
6324..6329
/note="splice donor sequence, gtagtgc, between exons 1 and
2 of Ca20C1.05"

misc_feature
6387..6406
/note="splice branch and acceptor sequence
TACTACAGGatgataatag between exons 1 and 2 of Ca20C1.05"
complement(6909..9200)
/gene="Ca20C1.06c"
complement(6909..9200)
/note="Ca20C1.06c, unknown, len: 763 aa, similar to S.
cerevisiae Q03388, D8035.28P (810 aa), fasta scores, opt:
339, E(1): 5.4e-29, (28.3% identity in 842 aa overlap)"
/codon_start=1

gene
CDS
complement(6909..9200)
/gene="Ca20C1.06c"
complement(6909..9200)
/note="Ca20C1.06c, unknown, len: 763 aa, similar to S.
cerevisiae Q03388, D8035.28P (810 aa), fasta scores, opt:
339, E(1): 5.4e-29, (28.3% identity in 842 aa overlap)"
/codon_start=1

[illegible]


```

misc_feature complement(<346..>1014)
/Note="similar to Saccharomyces cerevisiae ORF YER062c (
HOR2 : DL-glycerol phosphatase 1
similar to Saccharomyces cerevisiae ORF YII053w ( RHR2 :
DL-glycerol phosphatase 1"
/evidence=not_experimental
BASE COUNT      350 a      189 c      179 g      297 t      2 others
ORIGIN

alignment_scores:
  Quality: 333.50      Length: 198
  Ratio: 2.508      Gaps: 4
  Percent Similarity: 67.172      Percent Identity: 39.899

alignment_block:
US-09-558-284-2 x CNS07CWN/rev ..

Align seg 1/1 to reverse of: CNS07CWN from: 1 to: 1017

37 ValasProSerGluLeuPheLysHisSerHisGlyAlaArgThrGlnG1 53
|||||.....|.....|.....|.....|.....|.....|.....|
963 GTTATCTCTGAGAAATATTAGAACCTCCATGCTGCTCTACCTACGA 914
|||||.....|.....|.....|.....|.....|.....|.....|
53 uValleuArgArgPhePheProLysLeuAspSphThraspAnlyGlyV 70
|||||.....|.....|.....|.....|.....|.....|.....|
913 TGTATGCGCCAAATGCTCCAGCA.....GATCGTATGTCGACACAA 870
|||||.....|.....|.....|.....|.....|.....|.....|
70 alLeuAlaLeuGluLysAspIleAlaHisSerTYrLeuAspThrValSer 86
|||||.....|.....|.....|.....|.....|.....|.....|
869 TTACTGCTGGGAAGTCCATTCAGATCTTTGTTGTAATACCCCAA 820
|||||.....|.....|.....|.....|.....|.....|.....|
87 LeuIleProGlyAlaGluAsnLeuLeuSerLeuAspValAspThrG1 103
|||||.....|.....|.....|.....|.....|.....|.....|
819 CCAATTCGGGCTGCTGTTGAATTCGTTAAATCTTTGATTAATTCCTCA 770
|||||.....|.....|.....|.....|.....|.....|.....|
103 uThGlnLysLysLeuProGluArgLysTrpAlaIleValThrSerGlyS 120
|||||.....|.....|.....|.....|.....|.....|.....|
769 AGAACAAACCGAAACGCTCAACAAAGATGCGCTGCTGCTGCTGCTGA 720
|||||.....|.....|.....|.....|.....|.....|.....|
120 erProTYrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGly 136
|||||.....|.....|.....|.....|.....|.....|.....|
719 CTTTACCATTTAGCCCAATGCTTGAAA.....TTGTTAACCACTGAA 676
|||||.....|.....|.....|.....|.....|.....|.....|
137 LysProLysValPheIleThrGlyPheAspValLysAsnGlyLysProAs 153
|||||.....|.....|.....|.....|.....|.....|.....|
675 AAACGACATTTTTCATCTACTGCGCAAAAGTCACTAAAGTAAACCA 626
|||||.....|.....|.....|.....|.....|.....|.....|
153 pProGluGlyTYrSerArgAlaArgAspLeuLeuArgGlnAspLeuGlnL 170
|||||.....|.....|.....|.....|.....|.....|.....|
625 TCCACAAAGGTATACCAAGCGCTAGAGATCTTTG..... 592
|||||.....|.....|.....|.....|.....|.....|.....|
170 eutThrGlyLysGlnAspLeuLysTYr.....ValValPheGluAspAla 184
|||||.....|.....|.....|.....|.....|.....|.....|
591 .....GGTTACACAGATGCTCACAAGTGGCTGTTTTCGACATGCT 547
|||||.....|.....|.....|.....|.....|.....|.....|
185 ProValGlyIleLysAlaGlyLysAlaIleMetGlyAlaIleThrValGlyI 201
|||||.....|.....|.....|.....|.....|.....|.....|
546 CCAAGTGTATCACTGCTGTTAAAGTCTGCGCATTTGTCATTTGGTAT 497
|||||.....|.....|.....|.....|.....|.....|.....|
201 eThrSerSerTYrAspLysSerValLeuPheAspAlaGlyAlaAspTYrV 218
|||||.....|.....|.....|.....|.....|.....|.....|
496 CTGTTTACTATGACCAAGAAAGTTAGAAATCTGCTGCCAACATTTG 447
|||||.....|.....|.....|.....|.....|.....|.....|
218 alValLysAspLeuThrGlnValSerValValLysAsnAsnGlu 232
|||||.....|.....|.....|.....|.....|.....|.....|
446 TCGCGCAAGACTTGTCAAGCTTCAAGATTGATCTTACACAAA 403
|||||.....|.....|.....|.....|.....|.....|.....|

seq_name: gb_p14:YSPRHR2C
seq_documentation_block: 925 bp mRNA PLN 10-FEB-1999
LOCUS YSPRHR2C
DEFINITION Yeast Rhr2 mRNA, complete cds.

```

```

ACCESSION D50471
VERSION D50471.1 GI:1236749
KEYWORDS Rhr2.
SOURCE Saccharomyces cerevisiae (strain:RS16) cDNA to mRNA.
ORGANISM Saccharomyces cerevisiae
REFERENCE 1 (bases 1 to 925)
AUTHORS Hirayama,T.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-1995) to the DDBJ/EMBL/GenBank databases. Takashi
Hirayama, T. The Institute of Physical and Chemical Research (RIKEN);
Koyadai 3-1-1, Tsukuba, Ibaraki 305, Japan
(E-mail:hirayama@rctsl.riken.go.jp, Tel:0298-36-4359,
Fax:0298-36-9060)
REFERENCE 2 (bases 1 to 925)
AUTHORS Hirayama,T., Maeda,T., Saito,H. and Shinozaki,K.
TITLE Cloning and characterization of seven cDNAs for
hyperosmolarity-responsive (HOR) genes of Saccharomyces cerevisiae
JOURNAL Mol. Gen. Genet. 249 (2), 127-138 (1995)
MEDLINE 96086928
FEATURES
     source          location/Qualifiers
                    1..925
                    /organism="Saccharomyces cerevisiae"
                    /strain="RS16"
                    /db_xref="taxon:4932"
     gene            17..769
                    /gene="Rhr2"
     CDS             17..769
                    /gene="Rhr2"
                    /codon_start=1
                    /product="unknown"
                    /protein_id="BAA09060.1"
                    /db_xref="GI:1236750"
                    /translation="MPLTTKPLSLKINAALEFVDGTTIIISOPAIAAWRDPGKDPYF
                    DAELVNIHISGMWRYYDAIAKFAFDPADEEYVNNKEGSIPEKYGHSIEVGAVKLNLA
                    LNALPEKEMAVNATSGTRDMARKWEDILIKIRPEYFITANOVKQKPRPEYLGKGRNL
                    GFPINEDDPKSKVYVEDAPAGIAACAGACKIVGIATTFDIDFLKEKGCDDIIVKMH
                    ESTIRVGYNMETDEVELIFDDYLYAKDDLKMK"
BASE COUNT      298 a      191 c      176 g      260 t
ORIGIN

alignment_scores:
  Quality: 328.50      Length: 229
  Ratio: 2.313      Gaps: 8
  Percent Similarity: 62.009      Percent Identity: 37.555

alignment_block:
US-09-558-284-2 x YSPRHR2C ..

Align seg 1/1 to: YSPRHR2C from: 1 to: 925

10 LeuPheAspLeuAspGlyThrIleValSerThrThrValAlaIleGlyIu 26
|||||.....|.....|.....|.....|.....|.....|.....|
62 CATTCGATGTGACGGTACATCATCTCTCAACCGCATTCGCTGC 111
|||||.....|.....|.....|.....|.....|.....|.....|
26 salatrThrLysLeuGlyTYrGluTYrGly.....ValA 38
|||||.....|.....|.....|.....|.....|.....|.....|
112 TTTCGAGAGA.....GATTTGCGTAAAGACAAGCCTTACTTGG 149
|||||.....|.....|.....|.....|.....|.....|.....|
38 sPrProSerGluLeuPheLysHisSerHisGlyAlaArgThrGlnGluVal 54
|||||.....|.....|.....|.....|.....|.....|.....|
150 ATGCCGAAACAGCTTATTCACATCTCTCACGGTGGAGAACTTACGATGC 199
|||||.....|.....|.....|.....|.....|.....|.....|
55 LeuArgArgPhePheProLysLeuAspSphThraspAnlyGlyValle 71
|||||.....|.....|.....|.....|.....|.....|.....|
200 ATTGCCAAGTTCGTCACA.....GACTTTCCTGATGAGAAATACGTTAA 243
|||||.....|.....|.....|.....|.....|.....|.....|
71 uAlaLeuGluLysAspIleAlaHisSerTYrLeuAspThrValSerLeuI 88
|||||.....|.....|.....|.....|.....|.....|.....|
244 CAAGCTAGAAAGTGAATCCAGAAAAGTACGTTGACATCTCATGGAAG 293
|||||.....|.....|.....|.....|.....|.....|.....|

```

```

88 1eprg1yAlaGluAsnLeuLeuSerLeuAspValAspThrGluThr 104
   ::::::::::::::::::::
294 TTCAGGAGTCTCAAGTGTGTAATGCTTGAAGCC..... 331
105 GluLysLysLeuProGluArgLysTrpAlaIleValThrSerGlySerPr 121
   |||||:::|||||:::|||||:::|||||:::|||||:::
332 .....TTGCCAAGAGAAAAATGGCTGTCCGACCTCGTACCCG 372
121 oTrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysP 138
   ::|||:::|||||:::|||||:::|||||:::|||||:::
373 TGACATGGCCAGAAATGGTTCGAC...ATTTTGAAG...ATCAAGAGAC 416
138 rGlyValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
   ||:::|||||:::|||||:::|||||:::|||||:::|||||
417 CAGAACTACTATCAACGCCAATGATGTCAGCAAGTAAAGCTCAACCA 466
155 GluGlyTrpSerArgAlaArgAspLeuLeuArgGlnAspLeuGlnLeuTh 171
   |||:::|||||:::|||||:::|||||:::|||||:::
467 GAAACATACTTAAAGGTGAAACGGTTGGGTTTCCCAATTAAATGACCA 516
171 rGlyLysGlnAspLeuLysTrpValValPheGlnAspAlaProValGly 188
   ::|||:::|||||:::|||||:::|||||:::|||||:::
517 AGACCATTCCAATCTAAGTGTGCTTTTGAAGCAGCAGCTGGTA 566
188 1eLysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
   |||:::|||||:::|||||:::|||||:::|||||:::
567 TTGCTGCTGTAAAGCTGCTGCTGTAATGCTTGTATGCTACCACT 616
205 TyrAspLysSerValLeuPheAspAlaGlyAlaAspTrpValValCysAs 221
   ::|||:::|||||:::|||||:::|||||:::|||||
617 TTCGATTGTGACTCTTGAAGCAAGAGGCT.....TGTGA 651

221 pleuthrGlnValSerValValLysAsnGlnAsn 233
   |:::|||||:::|||||:::|||||:::|||||:::
652 CAYC.....ATTGTCAGAACCAAGCAATCT 676

seq_name: gb_p13:SCCEM1

seq_documentation_block:
LOCUS SCCEM1 2940 bp DNA PLN 15-OCT-1999
DEFINITION Saccharomyces cerevisiae CEM1 gene.
ACCESSION X73488.1 GI:403310
VERSION X73488.1 GI:403310
KEYWORDS beta-ketoacyl synthase; CEM1 gene.
SOURCE Saccharomyces cerevisiae.
ORGANISM Saccharomyces cerevisiae
          Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
          Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 2940)
AUTHORS Harrington, A., Herbert, C.J., Tung, B., Getz, G.S. and Slonimski, P.P.
TITLE Identification of a new nuclear gene (CEM1) encoding a protein
        homologous to a beta-keto-acyl synthase which is essential for
        mitochondrial respiration in Saccharomyces cerevisiae
JOURNAL Mol. Microbiol. 9 (3), 545-555 (1993)
MEDLINE 94018649
REFERENCE 2 (bases 1 to 2940)
AUTHORS Slonimski, P.P.
TITLE Direct Submission
JOURNAL Submitted (21-JUN-1993) P.P. Slonimski, Centre de Genetique
        Moleculaire, Centre National de la Recherche Scientif., Avenue de la
        Terrasse, Gif sur Yvette, F91198 Cedex, FRANCE
FEATURES
    source
    1..2940
        /organism="Saccharomyces cerevisiae"
        /strain="CM04"
        /db_xref="taxon:4932"
    gene
    1117..2932
        /gene="CEM1"
    TATA_signal
    1117..1121
        /gene="CEM1"
        /note="putative"
    TATA_signal
    1175..1180
        /gene="CEM1"

TATA_signal
1202..1205
/note="putative"
/gene="CEM1"
TATA_signal
1247..1251
/note="putative"
/gene="CEM1"
TATA_signal
1306..1309
/note="putative"
/gene="CEM1"
TATA_signal
1316..1320
/note="putative"
/gene="CEM1"
CDS
1453..2781
/note="putative"
/gene="CEM1"
/codon_start=1
/protein_id="CAB56180.1"
/db_xref="GI:6066682"
/translation="MSRRVLTGLGCVTPILGRSLSESWGNLSSKNGLPTITSIPNY
EDYKLRKSIPTSTTVCKIPENFQENSAIKLLETSODERTSSFKILALRTYEAL
HNAGLINPDITITMISLCNLDHFGLIGSGISTODIYOTSIQFNDKKRIAPYVPK
ILTNAGNVSIRKRNRLSHSVSTACTGNNSTDAFPTRLEKQDLCVAGASFTSL
HPLSLAFTRAKSLITTINGISRPFDIQSGFVUGBOGMIWBSLEHAKRANITSEL
VGYGLSDPACHITSPADNGAKRAIEMALMARLEPVDVYVNAHATSTLLGDAEC
LAVASALPGRSKSPILYISSNKAIGHLGAAGAVESIFITCSLKDMPHTYLDN
VUTLENNADKLHFIROKPIVGANPKYALCNISFGGVNTSLLEFKMEGS"
2884..2932
/gene="CEM1"
/note="consensus"
BASE COUNT 956 a 594 c 573 g 817 t
ORIGIN

alignment_scores:
    Quality: 328.50 Length: 237
    Ratio: 2.220 Gaps: 7
    Percent Similarity: 62.447 Percent Identity: 35.021

alignment_block:
US-09-558-284-2 x SCCEM1 ..
Align seq 1/1 to: SCCEM1 from: 1 to: 2940
2 AlaGluPheSerAlaAspLeuCysLeuPheAspLeuAspGlyThrIleVal 18
   ::::::::::::::::::::
418 TCCTCTTTGAAGTTAAAGCCGCTTGTGACGTCGACGATACATTAT 467
18 1SerThrThrValAlaAlaGluLysAlaTrpThrLysLeuCysTrpGlu 35
   ::|||:::|||||:::|||||:::|||||:::|||||
468 CATCTCTCAACCAAGCCATCTGCTGATCTGAGAG.....GATT 505
35 yrgLys.....ValAspProSerGluLeuPheLysHisSer 46
   ::|||:::|||||:::|||||:::|||||:::|||||
506 TCGTAAAGCAAAACCTTATTGATGCTGAACACGTTATCCAAAGTCTCG 555
47 HisGlyValArgThrGlnGluValLeuAlaArgPhePheProLysLeuAs 63
   |||||:::|||||:::|||||:::|||||:::|||||
556 CATGCTTGGAGAACGTTGATGCCATGCTAAGTTCCTCCA.....GA 599
63 pasPThrAspAsnLysGlyValLeuAlaLeuGluLysAspIleAlaHis 80
   |:::|||||:::|||||:::|||||:::|||||:::
600 CTTTGCCCAATGAAGAGATGTTAAACAATTAGAAGCTGAATTCGGGTCA 649
80 eryTrLeuAspThrValSerLeuIleLeuProGlyAlaGluAsnLeuLeu 96
   ::|||:::|||||:::|||||:::|||||:::|||||
650 AGTACGGTGAATAATCATTTGAAGTCCAGGTGAGTTAAGCTGTGCAAC 699
97 serLeuAspValAspThrGluThrGlnLysLysLeuProGluArgLysTr 113
   ::|||:::|||||:::|||||:::|||||:::|||||
700 GCCTTGAACGCT.....CTACCAAAAGGAATG 728
113 palAlaIleValThrSerGlySerProTyrLeuAlaPheSerTrpPheGlu 130
   |||||:::|||||:::|||||:::|||||:::|||||

```

```

729 GCCTGTGGCACTTCGGTACCCGTGATATGGCAAAAAATGTTGCAGC 778
130 hrl1leuLysAsnValGlySProlYsValPheIleThrGlyPheasp 146
    ::::: ::::: ||||| |||||::: |||
779 ATTCG.....GGAATCAGGACCAAGTACTTATTACCGCTATGAT 822
147 ValLysAsnGlySProlAspProGlnGlyTyrSerArgAlaArgAspLe 163
    |||||:::||||| ||||| ||| :::::|||||
823 GTCAACAGCGTAGCTCATCCAGAACCATATCTGAAGGCGCAGCAATCG 872
163 uleuArgGlnAspLeuGlnLeuThrGlyGlnAspLeuLysTyrValV 180
    ||| ::::: ::::: ||| |||
873 CTTAGGATATCCGATCAATGAGCAAGCCCTTCCAAATCTAAGGTAGTAG 922
180 alPheGluAspAlaProValGlyIleLysAlaGlyLysAlaMecGlyAla 196
    |||||:::||||| :::::||||| ||||| ||||| |||:::
923 TATTGGAAGACGCTCCAGCAGGTATTGCCGCCGAAAAAGCCCGCTTGT 972
197 IleThrValGlyIleThrSerSerTyrAspLysSerValLeuPheAsp 213
    ::|||:::|||||::: ||| :::::
973 AACATCATGTTGTTATTCACACTTTCGACTTGCATTGCTTAAGGAAAA 1022
213 agGlyAlaAspTyrValValGlyAspLeuThrGlnValSerValValLys 230
    ||| ::|||::: ::||| |||
1023 AGGC.....TGTGACATC.....ATTGTCANAA 1045
230 snAsnGluAsn 233
    ||:::|||||
1046 ACCCGAATCC 1056

```

seq_name: gb_p13:SC5610

seq_documentation_block:

LOCUS SC5610 37730 bp DNA PLN 20-AUG-1997

DEFINITION S.cerevisiae chromosome IX sequence derived from lambda clones

5610-5004.

238060.247047

238060.1 GI:557796

ATPase; delta element; hop1; meiosis specific protein; protein

transport; ribonucleotide-diphosphate reductase; ribosomal protein;

rlf3; sec6; small nuclear ribonucleoprotein; snpl; transfer

RNA-Set.

Saccharomyces cerevisiae.

Saccharomyces cerevisiae

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Saccharomyces.

1 (bases 1 to 37730)

Smith,V.

Unpublished

2 (bases 1 to 37730)

Barrell,B. and Rajandream,M.A.

Direct Submission

Submitted (02-SEP-1994) Saccharomyces cerevisiae chromosome IX

sequencing project, MRC Laboratory of Molecular Biology, Hills Rd,

Cambridge CB2 2QH, Sanger Centre, Hinxton Hall Hinxton, Cambridge

CB10 1HQ by Bart Barrell and Marie-Adele Rajandream. E-mail:

barrell@sanger.ac.uk

Notes:

All CDS over 100 codons have been analysed. CDS that are completely

overlapped and those that are overlapped by more than 50%

of their length by a larger CDS have been omitted from this

analysis.

Details of the omitted CDS are available on request. The more

significant matches with motifs in the PROSITE database are

also included but some of these may be fortuitous. The length in

codons and the calculated codon adaptation index (CAI)

is given for each CDS.

lambda 5610-5004 lie between cosmids 9877 and 9905, this sequence

is

contiguous with and 3' to that of cosmid 9877.

FEATURES

Source

1..37730

Location/Qualifiers

/organism="Saccharomyces cerevisiae"

/strain="AB972"
/db_xref="taxon:4932"
/chromosome="Chromosome IX"
/clone="lambda 5610-5004"
1..923
/gene="hop1"
<1..923
/gene="hop1"
/note="hop1"
/note="hop1, len: 306, CAI: 0.12, incomplete HOP1_YEAST
P20050 METOSIS SPECIFIC PROTEIN HOP1"

gene

CDS

/codon_start=3
/protein_id="CAAB6151.1"
/db_xref="GI:557797"
/db_xref="SWISS-PROT:P20050"
/translation="TNNVIGNLINSQASIOPTQPVSNPNPTGICSECGLEVPKAT
VKTCKSCRKLTGICYNGLHSIRKCPICIGPSLDTPKMSFODLPMIRKFRRLV
RKKGFPASTITELIDFTINVEDONNEKKEVAAALVFFLIDFLCLDNGKPSQOTIRY
VTSSVLVDKGVIPNTRKQLNVNHEKXMTTSSPKAESFYQEVLPNSRKVESWLD
DITNLKRVSEALSPSTLDELNSLPPQDPTISQAKRRRYDLDEYLEEDKSSVN
DTTKADFDESVPAKIRKISVSKRTLSNM"
complement(1164..2498)
/note="orf, len: 444, CAI: 0.12"
/codon_start=1
/protein_id="CAAB6152.1"
/db_xref="GI:557798"
/db_xref="SWISS-PROT:P40512"
/translation="MFGAKPLILERIGHLISINKEEERKRMANOGISTYLOEVC
TSPYLEILEESGLRPLVSLNSQLGKPHPSILGGPDENIARDIISHNPMALPOME
SEEVPLTRKVOHEKTIQISLCKNRKGIIEIYNNKNTIOGKRFNPMNSMKDRIS
HEVVDSDFSFLRIOMLCVSTFLOERTFDCTKFTFMSTSEPLTLKVLSEHIDKMF
ISKEEFIMVNVNISVLISPLNDYDEIYLSDKQFQMPPLVNCLELINTPNKFE
KIWGEINKICMEKSLFLEPSMSSAAVIMRCKIYFYRLRSKRLQFSYLSSTGIDLE
DIKEELTKLIISQNLFEIDGVIFHEDSILQSIYNEISRNGTMIINEVIDKLNEVT
DIKDIQGNPLWMSGNNTATITINNESSDMDIDEVNDRSDISDSGLPEEC"
complement(2776..3576)
/note="orf, len: 266, CAI: 0.21"
/codon_start=1
/protein_id="CAAB6153.1"
/db_xref="GI:557799"
/db_xref="SWISS-PROT:P40513"
/translation="MFLRSYNRAVTSILTTTPRAYVSSKRVFTYANSKRCTPAI
MRQETQVRGDIILQSELEKTEKTEPSTSIDSPNDLNTKYSFSLVEIPGKNEIVAR
TESGETVHVEFDVAQIANLPYNNAMDENFDNDGJNEDFDALSDNFANVVISKE
SASEPAVSPELLNLOESFVYDSATPYSDNALQSAEAETIRELYRGLAEKVEK
EILQESLPAVLESRGVNEIASVSEKNEKNEYSMEKKKKFFH"
complement(4057..4461,4871..4873)
/note="spliced, ribosomal protein, len: 135, CAI: 0.76,
similar to RS24_HUMAN P16632 40S RIBOSOMAL PROTEIN S24
(S19)"
/codon_start=1
/protein_id="CAAB6154.1"
/db_xref="GI:557800"
/db_xref="SWISS-PROT:P26782"
/translation="MSDAVITIRTRKVISNPLARKQFVVDVLHPNRAVSKDELREKL
AEVQAKKAVSVFGRTOFGGKSVGFLVNSVBAKFEPTRLVRLVGLAEKVEK
ASFOQRKOKKRNBDKIFCTGKRLAKKVAARNAD"
complement(4462..4464)
/note="splice acceptor sequence TAG"
complement(4518..4524)
/note="splice branch sequence TACTAAC"
complement(4865..4870)
/note="splice donor sequence GTATGT"
complement(5561..7978)
/gene="sec6"
complement(5561..7978)
/gene="sec6"
complement(5561..7978)
/note="sec6, len: 805, CAI: 0.14, conflict with SEC6_YEAST
P33844 PROTEIN TRANSPORT PROTEIN SEC6"
/codon_start=1
/protein_id="CAAB6155.1"
/db_xref="GI:557801"
/db_xref="SWISS-PROT:P33844"
/translation="MSSDPLQGVCDLIKGLDLSLEVRVDIKKQLKEKSVVEYOLANKES

CDS

CDS

misc_feature

misc_feature

misc_feature

misc_feature

gene

gene

CDS

```

DYKGEVEESLKLNLKSNVTSIKOINEVKNLGNDRNPAINRVDILFRATKIVETV
NTSSYDRITNFVLMHEIERLIVAELEADLETCPPHLLLEHILLTSARPODEBV
VNAKRETEAORPTAKLEFRLSGIISKEFKLLDGLTYDIVEVARAOISLAIRLEYV
DLBERDLEIEAIRNTIKKELEIEKSSIKLPLNSNARLDDETPKVEIPLNGLY
QEMSGTISTRTAPRGYKHEHLLINGINNSISEMFKREKREYEDOKDVKDNDMJE
LIVIEHIANCCPWHNIEVEYFDQYKELSHLDIVLESEPTIILIDLEAFKFO
DLKODPFTKSEKSVIGDKETELFKDYLNIYKMTVEIGNLEKAEFDVLEPST
PHSDSGLLFLDGTCTCFOMTEOVENAGNOAKILVGVRESDILTRKOKNMIS
KISEIKKOINYNNKHDDIPESITPEDEKGLVEVILAVSDOKKADYAVATISKY
GLVSKYVKQJTNHLEGTLDSPFAVQSSIGLITLMDPDLRKPQOIEKSTWGS
OAOQJADTLDLTDIKPQMSVLEVFVQNSVIGETILKPLATSEHSEKNNKFL
EAMKRPDELTVQVLEKVLVDGNSKDTLITONFTVMEFMDLSCPTIDSLDITWQYLE
VYMDSRIDLVGLKCRKDVSSSEKRIVOQATEMLHEYRRMMEANGVDRBETLMRRF
VLEPFO"
misc_feature
complement(6689..6930)
/ gene="sec6"
/ note="HSA67F071 213703 H. sapientis partial cDNA sequence
clone 67F07 strand (-) single read"
complement(8228..10264)
/ note="orf, len: 678, CAI:0.11"
/ codon_start=1
/ protein_id="CA86156.1"
/ db_xref="GI:557802"
/ db_xref="SWISS-PROT:P40514"
/ translation="MGVHEDDANTTWEATDPGVSSDDCGHRTESIOLQNFSTDM
ESMLDEBGRNSKSKMLLKRKHPIOKFLERVMNVPPEDEPSPFKRMGLKIID
FPQPTFKTIRISKILRLILLYCCFMRIEPLILYPLIPYHPHMDGSKRPIIS
LSCNSYLMENGCNCKLNKNGPLDNKEVIRTRPACIDRCGWYTSALVAGNRVKY
TCYETGGALSEEDPKNVYSPYKSDSPCCASAVAGYISPYSGCTVSMQGANSE
PSKGMVNTGFSVAFNSFPQSGYFRDIOGGLSCQYRAVAVALNMLFGPIEYLY
DSLYGWLNTIVGWTLVLSLDPPLTDADHPASVEYELFSGVTLPLCLCAALVYVK
SAVKTLENGSPILAKVILMPTFMGLISNNYFDEPLDRLTWDILKROCAALVYVK
IAATILTCAVIOAVSLKMSGRPKYKFIYIOGGLIAGSLPGINLIHHYILGSLI
VPGCATRGSSAVYLFQGLVGLILSGVAREASIVETDITALLRGGAGSLKRPILIDN
DDONHSLSMHNAIDPVIDOJGNIDGFSLLADVEYVGNKMTVSIDVLRNENPLAQ
MMDDLAENSNTIDILYKARASVSPNKGDDYTNAGVLOWPENGMOKPEBGVS"
gene
complement(10603..13260)
/ gene="ftr3"
complement(10603..13260)
/ gene="ftr3"
/ note="ftr3, len: 885, CAI: 0.29, conflict with R13_YEAST
P21672 RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE"
/ codon_start=1
/ protein_id="CA86157.1"
/ db_xref="GI:557803"
/ db_xref="SWISS-PROT:P21672"
/ translation="MTSKNSSSNKSNPTOMYVIRKDGKREPOFDKITSPTLSYG
LDPRIKAVKTORILISGVSGVTTELDNLAARCAVNTTVHEDYATLARIALSNL
HKOTTKQSKVIDLHDMINATGKHAPISDEITNYVMKDTLNSAIVDRDFOYT
YFGKTLERSLTLKNGEVARPOHLYMVAALGHSIDIESVLTYNLMSLRYSTAS
PTLENAQTPHPOMSSCFILAMKDSIEGIDYTLKCAMISKTAGVGLINNIISTGS
YIACTNGTISNGLIPMIRVFNTARYVDQGGKRGAGALFLEPHADLFDNLRKTH
GKEIRARDLEPALMIDPLEPKRVQEDGPTLFSPSAPGLDVGWGEFEELVRYER
EGSRKTIKAOKLWAIILAOETGCTPFWYKDCANRKTNOONCTGKISSNCCGIVEY
SSPETAVCNINASTALPAPFVSPDCTASTASNEERLHEIKAVITTHNRYLDNRYV
PEANSNMKRRPILAVQGLADTYMLRLPPESEAEOTLNQOILETLYHATVNSCEL
AOKGKYSTEGSPASKGILQFDMNNAKPGEMWELRKDYVHGGHNSITJVMAPMT
ASTSOILGVNECEPVTNMYSRVLSGEFOVWYPLRLDVLGIDMDSMVKOYLIO
NGSIOGLPNVPOELKELYKTWEISOKTIIIMADRAKITVDSHSLPLPOAPBMGI
TSMHYGMKKGLGTGYLRTQOASAAIOFTIDEVADDOATHTHASYSELDPRYVK
GTRSEQKASALTESDNEDKASPVSEOSVSSANSNKLSDSVAPVAFETTKED
SDEKCDIYNEKVIACATPPPEACSGSG"
misc_feature
complement(11392..11472)
/ gene="ftr3"
/ note="Ribonucleotide reductase large subunit signature"
complement(11410..11472)
/ gene="ftr3"
/ note="Multi-copper oxidases signature 1"
complement(11812..11479)
/ note="orf, len: 155, CAI: 0.11"
/ codon_start=1
/ protein_id="CA86158.1"
/ db_xref="GI:557804"

```

```

CDS
/db_xref="SWISS-PROT:P40515"
/translation="MTKVDFPPTLKDAAEPLVPOQLLELRQVSEGGPTATIOSRFN
YAMGLIKSTVDVDERLGVKILITDIKKEASRRRECLYLTIGCYKLGESYMAKRYVD
LFEHRRNKVOVALKSMWEKIDKFORETLGVAVAGVLGAVAVASFRLRNRR"
14447..15220
/ note="orf, len: 257, CAI: 0.13"
/ codon_start=1
/ protein_id="CA86159.1"
/ db_xref="GI:557805"
/ translation="MKRESEKSSSALKNGIMERTPEKVVONOGTADLSTKIGTK
YWDELALTELEFRNPQDGTDCWPSDSDAEOKMIDPLVDNIGAYRISANSYVDLCT
GNGHMLPELHOTEFRQGRVGDIDSESKIASINAEATGVNFIISFOADTIFSGDMKP
GKYDIVLDKGLDILSGMKINGKLDVYVAGVERILKIDGIFLITSCNFTODEL

alignment_scores:
Quality: 328.50 Length: 229
Ratio: 2.313 Gaps: 8
Percent Similarity: 62.009 Percent Identity: 37.555

alignment_block:
US-09-558-284-2 x SC5610
Align seg 1/1 to: SC5610 from: 1 to: 3730

10 Leupheaspleuaspglythrillevalserthrvalalaalaglu 26
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27665 CTATTCGATGCTTGCAGTACATCATCTCTCAACACGACCATTCGCTGC 27714
26 salatprthrlyleucysyrgtutrgly ..... ValA 38
|||||
27715 TTTCCTGGACA ..... GATTTCGATTAAGACCAAGCTTACTTCG 27752
38 spProsergluleupheulysHisSerHisglYAlaAlythrIngluVal 54
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27753 ATGCCCAACACGTTATTCACATCTCTCACGCTTGGAGACTTACGATGCGC 27802
55 leuArgAArgpheheprolysleuasparthrAspAsnlyslYAla 71
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27803 ATTGCCAAGTCTGCTCCA ..... GACTTGCCTGATGAGAAATACGTTAA 27846
71 uAlaAeugluLysAspIleAlaHisSerThrLeuAspThrValSerLeu 88
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27847 CAAGCTAGAGGAGTGAATGCCAGAAACGATGAGTACACCTCCTCAGTGAAG 27896
88 IeprocIylAaGluAsnLeuLeuSerLeuAspValAspThrIuThr 104
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27897 TTCACAGTCTCTGTCAAGTTGTGTAAATGCTTGAACGCC ..... 27934
105 GluLysLysLeuprogIuarglYstrPalalaIleValThrSerGlySerPr 121
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27935 ..... TTCCCAAGCAAAAAATGGCGTGGCGCACCTCTGTGTAACCG 27975
121 ofYrleuAlaPheSerTrrpIhegluThrIleleuLysAsnValGlyLysP 138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27976 TCACATGGCCCAAGAAATGTTGAC...ATTTTGAAG...ATCAGAGAGC 28019
138 polyValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
28020 CAGAACTACTTCATCCACCGCAATGATGTCACCAAGGTAAGCTCACCCA 28069
155 GluGlyTyrSerArgAlaArgAspLeuLeuArgGlnAspLeuGluLeuThr 171
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
28070 GAAACCATCTTAAGGTAGAAACGTTGGCTTCCCAATTAATGAACA 28119
171 rGlyLysGlnAspLeuLysTyrValValPheGlnAspAlaProValGly 188
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
28120 AGACCCATCCCAATCTAAGGTTGTCTTTGAAGACCAACACAGCTGTGTA 28169
188 IeLysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
28170 TTGCTGCTGTGTAAGGCTGCTGCTGTAAATGCTGTAATTCACCACT 28219

```

205 tyraSplySerValleuPheAspAlaGlyAlaAspTyrValValCysAs 221
::: ||| : : : |||
28220 TTCGATTGGACTTCTTCAGAGAAAGGTT TGTGA 28254
221 pleuThrGlnValSerValValLysAsnGluAsn 233
::: ||| : : : |||
28255 CATC ATTGTCAAGACACGAAATCT 28279
seq_name: gb_pat2:AX085442
seq_documentation_block:
LOCUS AX085442 753 bp DNA PAT 09-MAR-2001
DEFINITION Sequence 55 from Patent WO0112833.
ACCESSION AX085442
VERSION AX085442.1 GI:13275496
KEYWORDS
SOURCE Saccharomyces cerevisiae.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE
AUTHORS 1 (bases 1 to 753)
TITLE Emplage, M., Haynie, S., Lafend, L., Pucci, J. and Whited, G.
Process for the biological production of 1,3-propanediol with high
titer
JOURNAL Patent: WO 0112833-A 55 22-FEB-2001;
F.I. DU PONT DE NEMOURS AND COMPANY (US) ; GENENCOR INTERNATIONAL,
INC. (US)
FEATURES
source 1..753
location/Qualifiers
BASE COUNT 225 a 162 c 173 g 193 t
ORIGIN
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
alignment_scores:
Quality: 327.50 Length: 229
Ratio: 2.323 Gaps: 7
Percent Similarity: 61.572 Percent Identity: 36.245
alignment_block:
US-09-558-284-2 x AX085442 ..
Align seq 1/1 to: AX085442 from: 1 to: 753
10 leuPheAspLeuAspGlyThrIleValSerThrThrValAlaIleGluVal 26
||||| : : : |||
46 TTGTTCCACGTCGACGCTACCATTAATCTCTCAACCAAGCATTTGCTGC 95
26 salATrPThrLysLeuGlyTyrGluTyrGly VALA 38
||| : : : |||
96 ATTCTGAGG GATTTCGTAGAGACAAACCTTATTCG 133
38 sPProSerGluLeuPheLysHisSerHisGlyAlaArgThrGlnGluVal 54
|| : : : |||
134 ATGCTGAACAGTTATCCAACTCTGCATGTGTGAGAACCTTTGATGCC 183
55 leuArgTrpPhePheProLysLeuAspAspThrAspAsnLysGlyValle 71
::: ||| : : : |||
184 ATTCTAAGTTGCTGCCA GACTTTGCCAATGAGAGATGTTAA 227
71 uAlaGluGlnLysAspIleAlaHisSerTyrLeuAspThrValSerLeu 88
||||| : : : |||
228 CAATTTAGAGCTGAATTCGCGTCAGTACGTCGCAAAATTCATTCAG 277
88 lePProGlyAlaGluAsnLeuLeuSerLeuAspValAspThrGluThr 104
::: ||| : : : |||
278 TCCAGGTCGACGTTAAGCTTGCAACGCTTGAACGCT 315
105 GlnLysLysLeuProGluArgLysTrpAlaIleValThrSerGlySerPr 121
||||| : : : |||
316 CTACCAAAAGAGAAATGGCTGTGGCAACTTCGGTACCG 356

121 oTyrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysP 138
::: ||| : : : |||
357 TGATATGGCACAATAATGTTGACATCTG GCAATCAGAGAC 400
138 rolYsValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
||||| : : : |||
401 CAAACTACTTCAATACCGCTAATGATGTCAACAGGTAAGCTCATCTCA 450
155 GluGlyTyrSerArgAlaArgAspLeuArgGlnAspLeuGlnLeuThr 171
||| ||| : : : |||
451 GAACCATATCTGAAGCGCAGGAATGCTTAGATATCCGATCAATGACGA 500
171 rGlyLysGlnAspLeuLysTyrValValPheGluAspAlaProValGlyI 188
::: ||| : : : |||
501 AGACCTTCCAAATCTAAGTAGTAGTATTGTAACAGCGTCACAGGTA 550
188 lElYsAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
||| ||| ||| ||| : : : |||
551 TTGCGCGCGGAAAGCCCGGTTGTGAATCATTTGTAATTCGCACTACT 600
205 tyraSplySerValleuPheAspAlaGlyAlaAspTyrValValCysAs 221
::: ||| : : : |||
601 TTCGACTTGGACTTCTTAAAGCAAAAGGC TGTGA 635
221 pleuThrGlnValSerValValLysAsnGluAsn 233
::: ||| : : : |||
636 CATC ATTGTCAAAACACCAATCC 660
seq_name: gb_pat4:YSCHOR2A
seq_documentation_block:
LOCUS YSCHOR2A 969 bp mRNA PLN 10-FEB-1999
DEFINITION Yeast Hor2 mRNA, complete cds.
ACCESSION D50469
VERSION D50469.1 GI:1236745
KEYWORDS
SOURCE Hor2.
ORGANISM Saccharomyces cerevisiae (strain:RS16) cDNA to mRNA.
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE
AUTHORS 1 (bases 1 to 969)
TITLE Hirayama, T.
JOURNAL Direct Submission
Submitted (01-MAY-1995) to the DDBJ/EMBL/Genbank databases. Takashi
Hirayama, T., Maeda, T., Saito, H. and Shinozaki, K.
Cloning and characterization of seven cDNAs for
hyperosmolarity-responsive (HOR) genes of Saccharomyces cerevisiae
Mol. Gen. Genet. 249 (2), 127-138 (1995)
FEATURES
source 96086928
location/Qualifiers
1..969
/organism="Saccharomyces cerevisiae"
/strain="RS16"
/db_xref="taxon:4932"
47..799
/gene="Hor2"
47..799
/gene="Hor2"
/codon_start=1
/product="unknown"
/protein_id="BA09058.1"
/db_xref="GI:1236746"
/translation="MGUTTKPLSLKYNALFDVDCGIIISOPAIAMFWRDFKDKPYF
DAEHVIOVSHGWRTPDAIAKFAFDPAINEEVNKLAEAIPIKYEKSTIEPVKAGVAKLCA
LNLAPKEKXAVNAGSGTRDMAOKFHEHGLIRRPYFTTANVNRGGRKHPPEYLIKGRGL
GPTINODPSKSKSVVVFEDAPADAIAGKAAGCKIIGIATTFDLDPLKEKGCIDIYVNH
ESIRVGVGNAEIVDEVERIFDDYLYAADDLLKW"
BASE COUNT 318 a 196 g 260 t

ORIGIN

alignment_scores: Quality: 327.50 Length: 229
 Ratio: 2.323 Gaps: 7
 Percent Similarity: 61.572 Percent Identity: 36.245

alignment_block:
 US-09-558-284-2 x YSCHOR2A ..

Align seg 1/1 to: YSCHOR2A from: 1 to: 969

```

10 LeupheaspleuaspglythrIlevalSerthrThrValAlaIalagIuly 26
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
92 TTGTTGACGTCGACGTCATTCATCTCTCAACCGCATTCGTCGC 141
26 salATrThrLysLeucysTYGluTYrGLY.....ValA 38
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
142 ATTCTGGAGG.....GATTTCGTTAGGACAAACCTTATTTCG 179
38 sPProSerGluLeupheLysHisSerHisGlyAlaArgThrGlnGluVal 54
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
180 ATGCTGACACAGCTATCCAGCTCGCATGCTTGAGAACGTTGATGCC 229
55 LeuArgArgPhePheProLysLeuaspsrThrAspAsnLysGlyValLe 71
   ::| ::| ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
230 ATTGCTAGTTCGCTCA....GACTTTGCCAATGAAGAGATATCTTAA 273
71 uAlaleuGluLysAspIleAlaHisSerTYrLeuasprThrValSerLeu 88
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
274 CAATTTAGACAGTGAATTCGCGTCACAGTACGCTGAATAATCCATTGANG 323
88 lEProGlyAlaGluAsnLeuLeuSerLeuAspValAspThrGluThr 104
   ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
324 TCCGAGGTGCACTTAAGCTGTCAACGCTTTGAAGCT..... 361
105 GlnLysLysLeuProGluArgLysThrPalaIleValThrSerGlySerPr 121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
362 .....CTACCAAAAGAAATAGGCTGTGGCACTTCGCGTACCCG 402
121 oTyrlEuaAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysP 138
   ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
403 TCATATGGCACAAATAGTGTGACATCTG.....GGAATCAGGAGAC 446
138 rOlyValAlaPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
447 CAAGTACTTCACTACCGCTAATGATGCAACAGGGTAAGCCTCATCCA 496
155 GluGlyTYrSerArgAlaArgAspLeuLeuArgGlnAspLeuGlnLeuth 171
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
497 GAACCATATCTGAAGGCGACGAATGCTTAGATATCCGATCAATGAGCA 546
171 rGlyLysGlnAspLeuLysTYrValValPheGluAspAlaProValGlyI 188
   ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
547 ACACCTTCCAAATCTAAGTAGTATTGTAAGACGCTCCAGCAGGTA 596
188 lElysalIaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
597 TTGCCCGCCGAAAAAGCCCGGTTGTAAGATCATTTGTAATGGCCACTACT 646
205 TYrAspLysSerValLeupheAspAlaGlyAlaAspTYrValValCysAs 221
   ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
647 TTCGACTTGGACTTCTTAAGAAAAAGGC.....TGTTGA 681
221 pleuthrGlnValSerValValLysAsnAsnGluAsn 233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
682 CATC.....ATTGTCAAAAACACGCAATCC 706
  
```



```

9  ATGGCAGAAATTTTCAGCTGATCTATGCTTTTGAACCTAGATGATACAT 58
17  eValSerThrThrValAlaAlaGluLysAlaTrpThrLysLeuCysTyrG 34
59  AGTGAGTACAAACAGTGGCCGACAGAGAAACATGAGCAAGATTGTGTACG 108
34  lUTyG1ValAspProSerGluLeuPheLysHisSerHisGlyAlaArg 50
109  AATAGGGTGTGATCTCTCCGAGTATTTTAACCATTTTCATGTTGACAA 158
51  ThrGlnGluValLeuArgArgPhePheProLysLeuAspPThrAspAs 67
159  ACACAAGAGGTTTGAAGAGTTTTCCTAAATGGATGATACAGACAA 208
67  nLysGlyValLeuAlaLeuGluLysAspIleAlaHisSerTyrLeuAsp 84
209  TAAAGGTGCTTCTGCTAGAAAGAAATATGCCATAGTACTTGGA 258
84  hrValSerLeuIleProGlyAlaGluAsnLeuLeuSerLeuAspVal 100
259  CAGTAAGCCTTATCTCGTGCAGAGAACTTACTGTTATCGTTAGATGA 308
101  AspThrGluThrGlnLysLysLeuProGluArgLysTyrPalaIleValTh 117
309  GATACTGAGACTCAAAAAAGTTACTGTAAGGAAATGGGCTATCGTAC 358
117  rSerLysSerProTyrLeuAlaPheSerTrpPheGluThrIleLeuLys 134
359  CTTCTGGTTCATATTGGCATTTTCATGAGTTCGAGACAAATATGAAA 408
134  snValGlyLysProLysValPheIleThrGlyPheAspValLysAsnGly 150
409  ATGTTGGAAAGCCCAAAAGTTTCATTACTGGGTTTCACCTGAAAGACGG 458
151  LysProAspProGluGlyTyrSerArgAlaArgAspLeuLeuArgLins 167
459  AAGCTGTGATCCCGAGGGTATTCAAGAGCTCGTGAATTATTCGCTGA 508
167  PLeuGlnLeuThrGlyLysGlnAspLeuLysTyrValValPheGluAsp 184
509  TTTGCAATTAACTGGTAAACAGAGATCGAATATGTTGCTTCGAAAGATG 558
184  lAprOvalGlyIleLysAlaGlyLysAlaMetGlyValIleThrValGly 200
559  CACCGTGGGCAATAAAGCGCGCAAGCAATGGCGCATTACTGTGGGT 608
201  lIleThrSerSerTyrAspLysSerValLeuPheAspAlaGlyAlaAsp 217
609  ATTAACATCTCTGATGACAAAGCGTTTATTTGACGACGAGACAGATTA 658
217  rValValCysAspLeuThrGlnValSerValValLysAsnAsnGluAsnG 234
659  TGTAGTCTGTGATTTGCACACAGCTTCCGTGGTTAAGAAACATGAAGA 708
234  lYleValIleGlnValAsnAsnProLeuThrArgAla 246
709  GTATGTGATCCAGGTAACACACCTTTGACAAAGGCC 746

seq_name: /SIS1/gcdata/geneseq/geneseqn/NA2000.DAT:AAA13804
seq_documentation_block:
ID AAA13804 standard; DNA; 408 BP.
XX
XX
AC AAA13804;
XX
XX
DE 27-JUL-2000 (first entry)
XX
XX Yeast lacZ nucleotide sequence SEQ ID NO:3.
XX
XX Yeast: stress responsive gene; promoter; brewing; beer; wine; sake;
KM bread; oxidative stress; osmotic pressure; stress; glucose starvation;
XX ds.
XX

```

```

OS Saccharomyces cerevisiae.
FH Key Location/Qualifiers
FT CDS 81..407
FT /tag= a
FT /product= "lacZ"
FT /note= "no stop codon given"
XX
XX JP2000078977-A.
XX
XX 21-MAR-2000.
XX
XX 04-SEP-1998; 98JP-0251390.
XX
XX 04-SEP-1998; 98JP-0251390.
XX
XX (TAIF ) MARUHA CORP.
XX
XX MPI: 2000-285929/25.
XX
XX P-PSDB; AAY82568.
XX
XX A stress-responsive gene promoter
XX
XX Example 1; Page 9; 12pp; Japanese.
XX
XX The present invention describes a stress responsive gene promoter
XX isolated from Saccharomyces cerevisiae (Yeast). Also described in the
XX present invention are: (1) a promoter containing a DNA hybridising with
XX the above DNA under a stringent condition and having stress-responsive
XX promoter activity; (2) a gene expression cassette containing the above
XX promoter; (3) an expression vector containing the above gene expression
XX cassette; (4) a recombinant vector in which a gene encoding an optional
XX polypeptide is recombined to the above expression vector; (5) a
XX transformant containing the above recombinant vector; and (6) a method
XX for the preparation of the above polypeptide in which the above
XX transformant is cultured and the polypeptide is collected from the
XX resultant culture. Saccharomyces cerevisiae is used for the brewing of
XX beer, wine and sake and production of bread. The gene is responsive to
XX the stresses such as oxidative stress, osmotic pressure stress and
XX glucose starvation stress. The present sequence encodes lacZ, which is
XX used in an example from the present invention.
XX
XX Sequence 408 BP; 147 A; 65 C; 77 G; 119 T; 0 other:
XX
XX
XX alignment_scores:
XX Quality: 478.00 Length: 109
XX Ratio: 4.596 Gaps: 0
XX Percent Similarity: 95.413 Percent Identity: 82.569
XX
XX alignment_block:
XX US-09-558-284-2 x AAA13804 ..
XX
XX Align seg 1/1 to: AAA13804 from: 1 to: 408
XX
XX 1 MetAlaGluPheSerAlaAspLeuCysLeuPheAspLeuAspGlyThrI 17
XX ||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
XX 81 ATGCCACAATTTTCAGTACATCTTGTGCTTTTGACCTAGATGGAGCTAT 130
XX 17 eValSerThrThrValAlaAlaGluLysAlaTrpThrLysLeuCysTyrG 34
XX ||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
XX 131 TGTACACACACAACTGCAAGCGGAAAGTGCCTGGAATAATTATTCGCCG 180
XX 34 lUTyG1ValAspProSerGluLeuPheLysHisSerHisGlyAlaArg 50
XX ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
XX 181 AGCATGGGTTGATCCTGTTGACTTATTCAGGATTCCTCCATGGGCAAGA 230
XX 51 ThrGlnGluValLeuArgArgPhePheProLysLeuAspPThrAspAs 67
XX ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
XX 231 TCACAAGAAATGATGAAGAATTTTTCCAAATTTGGACATATCCGATTA 280
XX nLysGlyValLeuAlaLeuGluLysAspIleAlaHisSerTyrLeuAsp 84
XX ||| ::::::::::::::::::::::::::::::::::::::::::::::::::::

```

281 TAAAGCTGTTTCGGTTAGAAAAAGATATGCGAGATATTATTGACA 330
84 hrValSerLeuIleProGlyAlaGluAsnLeuLeuSerLeuAspVal 100
|||||
331 CAGTAAGCCTTATTCCTGCTGCGACAGAAATTTATTGTTATCTTAGATCTA 380
|||||
101 AspThrGluThrGlnIleLysLysLeuPro 109
|||||
381 GATACTGACGACTCAAAAAAGTTACTT 407

seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV42019

seq_documentation_block:
ID AAV42019 standard: DNA; 816 BP.
AC AAV42019;
XX
XX
DT 12-OCT-1998 (first entry)
XX
XX
DE Glycerol-3-phosphatase gpp1 gene.
XX
XX
KW Glycerol-3-phosphatase: gpp1 gene; 1,3-propanediol: ss.
XX
XX
OS Saccharomyces sp.
XX
XX
PN W09821339-A1.
XX
XX
PD 22-MAY-1998.
XX
XX
PE 10-NOV-1997; 97WO-US20292.
XX
XX
PR 13-NOV-1996; 96US-0030601.
XX
XX
PA (DUPO) DU PONT DE NEMOURS & CO E. I.
(GEMV) GENENOR INT INC.
XX
XX
PI Dias-Torres M, Gatenby AA, Haynie SL, Hsu AK, Lareau KD;
P1 Nagarajan V, Nair RV, Nakamura CE, Payne MS, Picatagalo SK;
P1 Trimbur DE, Whited GM;
XX
XX
DR WPI: 1998-297942/26.
DR P-PSDB: AAW30683.
XX
XX
PT Fermentative production of 1,3-propanediol - by single organism
PT containing cassette comprising specific genes, and capable of using
PT inexpensive carbon sources
XX
XX
PS Claim 10; Page 52-53; 95pp; English.
XX
XX
CC This DNA sequence includes the coding region of the gpp1 gene
CC encoding a cytosolic glycerol-3-phosphatase (see AAW30683) of
CC Saccharomyces sp. This enzyme catalyses the conversion of
CC glycerol-3-phosphate to glycerol. A claimed method for production
CC of 1,3-propanediol (I) comprises culturing a microorganism
CC transformed with a cassette containing at least 1 of the genes (see
CC AAV42012-21) for glycerol-3-phosphate dehydrogenase,
CC glycerol-3-phosphatase, glycerol dehydratase and 1,3-propanediol
CC oxidoreductase (see AAW30676-85). Any of these genes not present on
CC the cassette must be present endogenously. The enzyme sequences
CC may include substitutions, deletions and additions provided activity
CC is not altered. A single recombinant organism can now be used for
CC production of (I) from inexpensive C-sources (contrast use of
CC glycerol or dihydroacetone) without causing pollution. (I) is
CC a starting material for polyesters, polyurethanes and cyclic
CC compounds.
XX
XX
SQ Sequence 816 BP; 252 A; 178 C; 171 G; 215 T; 0 other;

alignment_scores:
Quality: 328.50 Length: 229
Ratio: 2.313 Gaps: 8
Percent Similarity: 62.009 Percent Identity: 37.555

alignment_block:
US-09-558-284-2 x AAV42019 ..
Align seg 1/1 to: AAV42019 from: 1 to: 816

10 LeuPheAspLeuAspGlyThrIleValSerThrThrValAlaAlaGly 26
|||||
109 CTATTGATGCTTGACGCTACCATCATCATCTCTCAACAGCATTGCTGC 158
|||||
26 AlaIleThrThrLysLeuLysCysTyrGlyTyGly.....ValA 38
|||||
159 TTTCTGGAGA.....GATTTCGTTAAAGACAGCCTTACTTTC 196
|||||
38 sPProserGluLeuPheLysHisSerHisGlyAlaIleThrGluVal 54
||
197 ATGCCGAAACAGTTATTCACATCTCTCAGGTTTGAGAACTTACGATGCC 246
|||||
55 LeuArgArgPhePheProLysLysAspThrAspAsnLysGlyValLe 71
::: |||||
247 ATTGCCAAGTTCCGCTCA.....GACTTTCGTGATGAGAAATACGTTAA 290
|||||
71 ValAlaLeuGluLysAspIleAlaHisSerTyrLeuAspThrValSerLeu 88
|||||
291 CAAGCTGAAAGGTGAATCCAGAAAGTGAAGTGAAGTCACTCGAC 340
|||||
88 IeProGlyAlaGluAsnLeuLeuSerLeuAspValAspThrGluThr 104
::: |||||
341 TTCAGGTGCTGTCAAGTTGTATGCTTTGAAGGCC..... 378
|||||
105 GlnLysLysLeuProGluArgLysTyrPalaIleValThrSerGlySer 121
|||||
379TTGCCAAGAAAGAAATGGCGTGTCCGACCTGTGTACCG 419
|||||
121 cTyrLeuAlaPheSerThrPheGluThrIleLeuLysAsnValGlyLysP 138
::: |||||
420 TGACATGCGCCAGAAATGTTGCAC...ATTTGAG...ATCAAGAGAC 463
|||||
138 rGlyValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
|||||
464 CAGATATTCTTATCACCAGCCCAATGATGTCAAGCAAGTAAAGCTCACCCA 513
|||||
155 GlnGlyTyrSerArgAlaArgAspLeuLeuArgGlnAspLeuGlnLeu 171
|| ||
514 GAACATACCTTAAAGGTAGAAAGCTTGGGTTTCCCAATTAAATGAAACA 563
|||||
171 rGlyLysGlnAspLeuLysTyrValValPheGluAspAlaProValGly 188
::: ||
564 AGACCCATCCAAATCTAAGGTTGTTCTTTGAAGACGACACAGCTGTA 613
|||||
188 IeLysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
||
614 TTGCTGCTGGTAAGCTGCTGCGCTGTAATAATCGTTGATATTGCTACACT 663
|||||
205 TyrAspLysSerValLeuPheAspAlaGlyAlaAspTyrValValCysAs 221
::: ||
664 TTGATTTTGACCTCTTGAAGAAAGGT.....TGTGA 698
|||||
221 pleuThrGlnValSerValValLysAsnAsnGluAsn 233
::: ||
699 CATC.....ATTGTCAGAAACACGAAATCT 723
|||||

seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV35740

seq_documentation_block:
ID AAV35740 standard: DNA; 816 BP.
AC AAV35740;
XX
XX
DT 28-SEP-1998 (first entry)
XX
XX
DE Klebsiella pneumoniae glycerol-3-phosphatase GPP1 gene.

KW GPPI gene; glycerol-3-phosphatase; production;
 KM 1,3-propanediol; recombinant; ss.
 XX Klebsiella pneumoniae.
 OS
 XX WO9821341-A2.
 PN
 XX 22-MAY-1998.
 PD
 XX 13-NOV-1997; 97WO-US20873.
 PE
 XX 13-NOV-1996; 96US-0030601.
 PR
 XX (GEMV) GENENCOR INT INC.
 PA
 XX Chase MW, Diaz-Torres M, Dunn-Coleman NS, Trimbur D;
 PI WPI: 1998-297944/26.
 DR
 XX
 XX New method for increasing production of 1,3-propanediol - comprises
 PT fermentation of inexpensive carbon sources by microorganism
 PT expressing dehydratase, used, e.g. to prolong half-life of enzyme
 XX
 PS Disclosure: Page 62; 133pp; English.
 XX
 XX The sequence is that of the GPPI gene which encodes cytosolic
 CC glycerol-3-phosphatase. It was used as part of a method of
 CC fermentative production of 1,3-propanediol (1,3-pd), using an organism
 CC comprising at least 1 gene encoding a dehydratase, is improved by
 CC inserting into the host a gene encoding protein X and culturing
 CC the transformant in presence of a carbon source (e.g. mono-, oligo-
 CC or poly-saccharide or 1C substrate) convertible to 1,3-pd.
 CC 1,3-pd is a starting material for polyesters, polyurethanes and
 CC cyclic compounds. 1,3-pd can now be produced by a single
 CC recombinant organism from inexpensive carbon sources such
 CC as glucose (rather than costly glycerol or dihydroxyacetone),
 CC rapidly and without causing pollution.
 CC
 SO Sequence 816 BP; 252 A; 178 C; 171 G; 215 T; 0 other;

alignment_scores:
 Quality: 328.50 Length: 229
 Ratio: 2.313 Gaps: 8
 Percent Similarity: 62.009 Percent Identity: 37.555

alignment_block:
 US-09-558-284-2 x AAV35740 ..

Align seg 1/1 to: AAV35740 from: 1 to: 816

```

10 LeupheaspLeuaspGlyThrIleValSerThrValAlaAlaGlu 26
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
109 CTAATCGATGTCGACGTCATCATCATCTCTCAACACGCACTTCTGC 158
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
26 salatrprhrlLysLeuGlyTyrGluTyrGly.....ValA 38
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
159 TTTCTGAGAGA.....GATTTCGTAAAGACAGACGCTTACTTCG 196
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
38 srrProSerGluLeuPheLysHisSerHisGlyAlaArgThrGlnGluVal 54
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
197 ATGGCCGAAACGTTATTCATCTCTCAAGCTTGGAGAACTTACGATGCC 246
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
55 LeuAtrGATrPhePheProLysLeuAspThrAspAsnLysGlyValLe 71
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
247 ATTGCCAGATTCGCTCA.....GACTTTCGTCAAGAAATACGTTAA 290
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
71 uAlaLeuGluLysAspIleAlaHisSerTyrLeuAspThrValSerLeu 88
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
291 CAAGCTAGAGAGTGAATCCAGAAAGTACGATGACACCTCATCGAAG 340
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
88 leProGlyAlaGluAsnLeuLeuSerLeuAspValAspThrGlnThr 104
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

341 TTCCAGGTGCTGTCAGATTCGTATATGCTTTGAACGCC..... 378
105 GlnLysLysLeuProGluArgLysTrpAlaIleValThrSerGlySerPr 121
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
379 .....TTGCCAAAGGAAAAATGGGCTGTGCCACCTCTGTACCCG 419
121 cTyrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysP 138
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
420 TGACATGAGCCAGAAAGTTCGAC...ATTTTGAAG...ATCAAGAGAC 463
138 rOLysValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
464 CAGAAATACCTCATCACCCCAATGATGTCAGACAGGTAAGCTTACCCA 513
155 GluGlyTyrSerArgAlaArgAspLeuLeuArgGlnAspLeuGlnLeuTh 171
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
514 GAACCATACTTAAGAGGTAGAAACGGTTGGCTTCCCAATTAATGAGACA 563
171 rGlyLysGlnAspLeuLysTyrValValPheGluAspAlaProValGly 188
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
564 AGACCCATCAATCTAAGGTTGTGCTTTTGAAGACGACACAGTGTGA 613
188 leLysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
614 TTGCTGCTGTGTAAGGCTGCTGCTGTAATAATGCTGTGTAATGCTACACT 663
205 TyrAspLysSerValLeuPheAspAlaGlyAlaAspTyrValValLysAs 221
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
664 TTCGATTTGGACTCTTGAAGGAAAGGTT.....TGTGA 698
221 pleuThrGlnValSerValValLysAsnAsnGluAsn 233
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
699 CATC.....ATTGTCAGAGAACCAACGAACTCT 723

```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV30825

seq_documentation_block:

ID AAV30825 standard; DNA; 816 BP.

XX AAV30825;

XX 14-SEP-1998 (first entry)

XX Saccharomyces cytosolic glycerol-3-phosphatase GPPI gene.

XX Glycerol-3-phosphatase; G3P; GPPI; RHR2; YIL053W; ss.

XX Saccharomyces sp.

XX WO9821340-A1.

XX 22-MAY-1998.

XX 10-NOV-1997; 97WO-US20293.

XX 13-NOV-1996; 96US-0030602.

XX (DUPLO) DU PONT DE NEMOURS & CO E I.

XX (GEMV) GENENCOR INT INC.

XX Butlhis BA, Gatenby AA, Haynie SL, Hsu AK, Lareau RD;

XX WPI: 1998-297943/26.

XX P-PSDB: AAW57327.

XX Fermentative production of glycerol using recombinant host -

XX containing genes for glycerol-3-phosphate dehydrogenase and/or

XX glycerol-3-phosphatase

XX Disclosure: Page 28; 57pp; English.

XX This Saccharomyces sp. gene, designated GPPI, RHR2 or YIL053W,

XX codes for a cytosolic glycerol-3-phosphatase (G3P, see AAW57327) that

CC catalyses the conversion of glycerol-3-phosphate to glycerol.
CC The invention provides recombinant organisms that express G3P and/or
CC glycerol-3-phosphate dehydrogenase (G3PDH) (see also AAW57324-32)
CC useful for the production of glycerol from a variety of C-sources.
CC A host cell is preferably transformed with a cassette containing
CC either a G3PDH gene and/or a G3P gene and then cultured in the
CC presence of a mono-, oligo-, polysaccharide or IC-substrate. The
CC glycerol obtained is used in cosmetics, liquid soaps, pharmaceuticals,
CC lubricants and antifreezes; its esters are used in the oil and fat
CC industries. The method produces glycerol rapidly and inexpensively
CC without generation of polluting by-products.

XX Sequence 816 BP; 252 A; 178 C; 171 G; 215 T; 0 other;

alignment_scores:

Quality: 328.50 Length: 229
Ratio: 2.313 Gaps: 8
Percent Similarity: 62.009 Percent Identity: 37.555

alignment_block:

US-09-558-284-2 x AAV30825

Align seg 1/1 to: AAV30825 from: 1 to: 816

10 LeuPheAspLeuAspGlyThrIleValSerThrValAlaIaGlyIu 26
109 CTATTTCATGTCGACGAGTACATCATCTCTCACACGACCATTCGTCG 158
26 SAlaIThrThrIleuScystrGluTyrGly.....ValA 38
159 TTTCTGGGGA.....GATTCGGTAAGACAAAGCCTTACTTCG 196
38 sPProSerGluLeuPheIleuHisSerHisGlyAlaArgThrGluIuVal 54
197 ATGCCCAACAGCTTATTCACATCTCTCACGCTTGAGAACACTTACATGCC 246
55 LeuArgArgPheProLysLeuAspThrAspAsnLysGlyValle 71
247 ATTGCCAAGTTCGCTCCA.....GACTTTCCTGATGAAGATACCTTAA 250
71 uAlaLeuGluLysAspIleAlaHisSerTyrLeuAspThrValSerLeu 88
291 CAAAGCTAGAAAGGTAAATCCCAAAAGTACGGTGAACCTCATCGAAG 340
88 IeProGlyAlaGluAsnLeuLeuSerLeuAspValAspThrGluThr 104
341 TTCAGGTGCTGTCAAGTTGTATGCTTGAACGCC..... 378
105 GluIleLysLeuProGluArgLysTrpAlaIleValThrSerGlySerPr 121
379TTGCCAAAGCAAAATGGCGCTGCCACCTGTGGTACCG 419
121 cTyrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysP 138
420 TGACATGCGCCAAAGAAATGCTTCGAC...ATTGGAAG...ATCAAGAGC 463
138 rOlusValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
464 CAGAAATCTTATCATCCGCAATGATGCAAGCAAGCAACCTCACCACA 513
155 GluIleLysSerArgAlaArgAspLeuAspGluAsnAspGluIleThr 171
514 GAACCATACCTAAAGGGTGAACCGTTGGTTCCCAATTAAATGAACA 563
171 rGlyLysGluAspLeuLysTyrValValPheGluAspAlaProValGlyT 188
564 AGACCCATCAAAATCTAAAGCTTGTGCTTTGGAACGCCACACCTGCTA 613
188 IeLysAlaGlyLysAlaMetGlyValaIleThrValGlyIleThrSerSer 204
614 TTGCTGCTGGTAAGGCTGCTGCTGTAATAATCGTTGCTATTGCTACCACT 663

205 TyrAspLysSerValLeuPheAspAlaGlyAlaAspTyrValValCysAs 221
664 TTGCATTTGGACTTCTTGAAGCAAAAGCT.....TGTCGA 698
221 IeThrGluIleValSerValValLysAsnGluAsn 233
699 CATC.....ATTGTCMAAGAACCCAGCAACT 723
seq_name: /SIDSL/gcdata/geneseq/geneseq/NA1999.DAT:AAx80614

seq_documentation_block:

ID AAX80614 standard; DNA; 816 BP.

AAx80614;

29-SEP-1999 (first entry)

GPPI gene.

GPPI: glycerol-3-phosphatase; glycerol; recombinant organism;

KW transformation; glycerol biosynthetic pathway; expression cassette;

KW 1-3-propanediol; pharmaceutical compound; antifreeze solution;

KW lubricant; polyurethane; cyclic compound; fat and oil industry;

polyester fiber; ss.

Saccharomyces sp.

Key Location/Qualifiers

FT CDS 1..816

FT /product= "glycerol-3-phosphatase"

FT /note= "cytosolic"

PN WO9928480-A1.

PD 10-JUN-1999.

PF 02-DEC-1998; 98WO-US25551.

PR 02-DEC-1997; 97US-0982783.

PA (DUPLO) DU PONT DE NEMOURS & CO E. I.

PA (GENEV) GENENCOR INT INC.

PI NaIr RV, Payne MS, Trimbur DE, Valle F;

DR WPI: 1999-385384/32.

DR P-PSDB: AAY26169.

XX Recombinant organisms containing G3PDH and or G3P phosphatase

PT Disclosure; Page 55; 84pp; English.

PS The present sequence is a GPPI gene encoding cytosolic glycerol-3

XX phosphatase which catalyses the conversion of glycerol-3-phosphate and

CC water to glycerol and inorganic phosphate. GPPI is alternatively

CC referred to as RHR2 or YII053W. This is used to produce glycerol from a

CC recombinant organism by transforming a suitable host cell with an

CC expression cassette comprising either one or both of the genes encoding

CC G3PDH and G3P, where the host cell has disruptions in either glycerol

CC kinase or glycerol dehydrogenase endogenous genes to prevent their

CC active expression. The transformed host cell is cultured with a carbon

CC source and glycerol is recovered. Compounds derived from the glycerol

CC biosynthetic pathway like 1,3-propanediol can also be produced. The

CC method provides a rapid, inexpensive and environment-friendly source of

CC glycerol. Glycerol is used in cosmetics, food, pharmaceuticals,

CC lubricants, anti-freeze solutions, fat and oil industry etc.. 1,3

CC -propanediol is used for the production of polyester fibers and the

CC manufacture of polyurethanes and cyclic compounds.

XX Sequence 816 BP; 252 A; 178 C; 171 G; 215 T; 0 other;

XX

[illegible]

```

55 LeuAArgPhePheProLysLeuAspSphThraspaNlysglyValle 71
   ::::::::::::::::::::
184 ATTGCTAAGTTCGCTCA.....GACTTTCCAATGAAGATGTTAA 227
71 uAlaLeuGluLysAspIleAlaHisSerTyrLeuAspThrValSerLeu 88
   ::::::::::::::::::::
228 CAAATTGAGAGCTGAATTCGGTCAAGTCAGGTAATAATCCATTGAAG 277
88 LeproGlyAlaGluAsnLeuLeuSerLeuAspValAspThrGluThr 104
   ::::::::::::::::::::
278 TCCAGGTGCAAGTTAAGCTGTGCAACGCTTGAACGCT..... 315
105 GluLysLysLeuProGluArgLysThrAlaIleValThrSerGlySerPr 121
   ::::::::::::::::::::
316 .....CTACCAAAAGAGAAATGGCTGTGGCAACTTCGGTACCG 356
121 cTyrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysP 138
   ::::::::::::::::::::
357 TGATATGGCACAAAATGTTGAGCATGTG.....GGAATCAGAGAC 400
138 rGlyValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
   ::::::::::::::::::::
401 CAAAGTCACTTACCGCTAATATGTCAAACAGCGTAACCTTCATCA 450
155 GluGlyTyrSerArgAlaArgAspLeuLeuArgLysAsnGluLeuTh 171
   ::::::::::::::::::::
451 GAACCATATCTGAAGGAGGAGATGCTTACGATATCCGATCATGAGCA 500
171 rGlyLysGluAspLeuLysTyrValValPheGluAspAlaProValGly 188
   ::::::::::::::::::::
501 AGACCTCTCAATCTAAGTAGTATTGTAAGACGCTCCACAGTA 550
188 IeLysAlaGlyLysAlaMerGlyAlaIleThrValGlyIleThrSerSer 204
   ::::::::::::::::::::
551 TTGGCGCGGAAAAAGCCGCTGTGAAGATCATGTGATTCGCCACTACT 600
205 TyrAspLysSerValLeuPheAspAlaGlyAlaAspTyrValValCysAs 221
   ::::::::::::::::::::
601 TTGCACTTGGACTTCTCTTAAGCAAAAAGGC.....TGTGA 635
221 pLeuThrGluValSerValLysAsnGluLys 233
   ::::::::::::::::::::
636 CATC.....ATGTCAAAACCAAGAAATCC 660
seq_name: /SIDSI/gcdata/geneseq/geneseqn/NA198.DAT.AAV35741
seq_documentation_block:
ID AAV35741 standard; DNA; 753 BP.
XX
AC AAV35741:
XX
DT 28-SEP-1998 (first entry)
XX
DE Klebsiella pneumoniae glycerol-3-phosphatase GPP2 gene.
XX
KM GPP2 gene; glycerol-3-phosphatase; production;
XX 1,3-propanediol; recombinant; ss.
XX
OS Klebsiella pneumoniae.
XX
XX
XX Key Location/Qualifiers
XX CDS 1..753
XX FT /*tag= a
XX FT /product= glycerol-3-phosphatase
XX
XX W09821341-A2.
XX
XX 22-MAY-1998.
XX
XX 13-NOV-1997; 97MO-US20873.
XX
XX 13-NOV-1996; 96US-0030601.

```

```

XX
PA (GENEV ) GENENCOR INT INC.
XX
PI Chase MW, Diaz-torres M, Dunn-coleman NS, Trimbur D;
XX
XX WPI; 1998-297944/26.
XX
XX P-PSDB; AAW60261.
XX
PT New method for increasing production of 1,3-propanediol - comprises
PT fermentation of inexpensive carbon sources by microorganism
PT expressing dehydratase, used, e.g. to prolong half-life of enzyme
XX
XX
XX Disclosure; Page 62-63; 133pp: English.
XX
XX The sequence is that of the GPP2 gene which encodes cytosolic
XX glycerol-3-phosphatase. It was used as part of a method of
XX fermentative production of 1,3-propanediol (1,3-pd), using an organism
XX comprising at least 1 gene encoding a dehydratase, is improved by
XX inserting into the host a gene encoding protein X and culturing
XX the transformant in presence of a carbon source (e.g. mono-, oligo-
XX or poly-saccharide or 1C substrate) convertible to 1,3-pd.
XX 1,3-pd is a starting material for polyesters, polyurethanes and
XX cyclic compounds. 1,3-pd can now be produced by a single
XX recombinant organism from inexpensive carbon sources such
XX as glucose (rather than costly glycerol or dihydroxyacetone),
XX rapidly and without causing pollution.
XX
XX Sequence 753 BP; 225 A; 162 C; 173 G; 193 T; 0 other.

```

```

alignment_scores:
Quality: 327.50 Length: 229
Ratio: 2.323 Gaps: 7
Percent Similarity: 61.572 Percent Identity: 36.245

```

```

alignment_block:
US-09-558-284-2 x AAV35741 ..

```

```

Align seg 1/1 to: AAV35741 from: 1 to: 753

```

```

10 LeuPheAspLeuAspGlyThrIleValSerThrThrValAlaIleGly 26
   ::::::::::::::::::::
46 TTGTCGACGTGACGATCATCTATCATCTCAACAGCATTTGCTGC 95
26 sAlaTrpThrLysLeuCysTyrGluTyrGly.....ValA 38
   ::::::::::::::::::::
96 ATTCTGACAG.....GATTCGTTAAGACAAACCTTATTTCG 133
38 sProSerGluLeuPheLysHisSerHisGlyAlaArgThrGluLysVal 54
   ::::::::::::::::::::
134 ATGCTGAACAGCTTATCCAGTCGCACTGCGTGTGGAGACGTTGATGCC 183
55 LeuAArgPhePheProLysLeuAspSphThraspaNlysglyValle 71
   ::::::::::::::::::::
184 ATTGCTAAGTTCGCTCA.....GACTTTCCAATGAAGATGTTAA 227
71 uAlaLeuGluLysAspIleAlaHisSerTyrLeuAspThrValSerLeu 88
   ::::::::::::::::::::
228 CAAATTGAGAGCTGAATTCGGTCAAGTCAGGTAATAATCCATTGAAG 277
88 LeproGlyAlaGluAsnLeuLeuSerLeuAspValAspThrGluThr 104
   ::::::::::::::::::::
278 TCCAGGTGCAAGTTAAGCTGTGCAACGCTTGAACGCT..... 315
105 GluLysLysLeuProGluArgLysThrAlaIleValThrSerGlySerPr 121
   ::::::::::::::::::::
316 .....CTACCAAAAGAGAAATGGCTGTGGCAACTTCGGTACCG 356
121 cTyrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysP 138
   ::::::::::::::::::::
357 TGATATGGCACAAAATGTTGAGCATGTG.....GGAATCAGAGAC 400
138 rGlyValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154

```



```

seq_documentation_block:
ID   AAX80615 standard; DNA: 753 BP.
XX
AC   AAX80615.
XX
DT   29-SEP-1999 (first entry)
XX
DE   GPP2 gene.
XX
KW   GPP2; glycerol-3-phosphatase; glycerol; recombinant organism;
KW   transformation; glycerol biosynthetic pathway; expression cassette;
KW   1-3 propanediol; pharmaceutical compound; antifreeze solution;
KW   lubricant; polyurethane; cyclic compound; fat and oil industry;
KW   polyester fiber; EC 3.1.3.-; ss.
XX
OS   Saccharomyces cerevisiae.
XX
FH   key
FT   CDS
FT   1..753
FT   /tag= a
FT   /product= "glycerol-3-phosphatase"
FT   /note= "cytosolic"
XX
PN   MO9928480-A1.
XX
PD   10-JUN-1999.
XX
PF   02-DEC-1998; 98MO-US25551.
XX
PR   02-DEC-1997; 97US-0982783.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E. I.
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Nair RV, Payne MS, Trimbur DE, Valle F;
XX
DR   WP1: 1999-385384/32.
DR   P-PDB: AAY26170.
XX
PT   Recombinant organisms containing G3PDH and or G3P phosphatase
XX
PS   Disclosure: Pages 55-56; 84pp; English.
XX
CC   The present sequence is a Gpp2 gene encoding cytosolic glycerol-3-
CC   phosphatase which catalyses the conversion of glycerol-3-phosphate and
CC   water to glycerol and inorganic phosphate. GPP2 is alternatively
CC   referred to as HOR2 or YER062C. This is used to produce glycerol from a
CC   recombinant organism by transforming a suitable host cell with an
CC   expression cassette comprising either one or both of the genes encoding
CC   G3PDH and G3P, where the host cell has disruptions in either glycerol
CC   kinase or glycerol dehydrogenase endogenous genes to prevent their
CC   active expression. The transformed host cell is cultured with a carbon
CC   source and glycerol is recovered. Compounds derived from the glycerol
CC   biosynthetic pathway like 1,3-propanediol can also be produced. The
CC   method provides a rapid, inexpensive and environment-friendly source of
CC   glycerol. Glycerol is used in cosmetics, food, pharmaceuticals,
CC   lubricants, anti-freeze solutions, fat and oil industry etc.. 1,3
CC   -propanediol is used for the production of polyester fibers and the
CC   manufacture of polyurethanes and cyclic compounds.
XX
SQ   Sequence 753 BP: 225 A; 162 C; 173 G; 193 T; 0 other;

alignment_scores:
Quality: 327.50      Length: 229
Ratio: 2.323      Gaps: 7
Percent Similarity: 61.572      Percent Identity: 36.245

alignment_block:
US-09-558-284-2 x AAX80615 ..
Align seg 1/1 to: AAX80615 from: 1 to: 753

```

```

10  LeuPheAspLeuAspGlyThrIleValSerThrThrValAlaIleGly 26
    |||||.....:|||||.....:
46  TTGTCGACGTCGACGTCGATTCATCTCTCAACCGACCATTCGTCG 95
    .....:|||||.....:
26  sAlatrPhrIlyLeuCysTyrGluTyrGly.....ValA 38
    |||
96  ATTCTGAGAG.....GATTCGCTAAGAGCAAAACCTTATTTCG 133
38  sPProSerGluLeuPheLysHisSerHisGlyAlaArgThrGlnGluVal 54
    || .....:|||||.....:
134  ATGCTGCAACACGCTTATCCAGTCGTCGATGTTGGAGAACGCTTTATGCG 183
55  LeuArgArgPhePheProLysLeuAspThrAspAsnLysGlyValle 71
184  ATTGCTAAGTTCGCTCA.....GACTTGGCAATGAAGATATGTTAA 227
71  uAlaLeuGluLysAspIleAlaHisSerTyrLeuAspThrValSerLeu 88
    |||||.....:
228  CAAATTAGAACGCTGAATTCGCGTCAGTCAGTCGTAATAATCATTTAG 277
88  leProGlyAlaGluAsnLeuLeuSerLeuAspValAspThrGluThr 104
    .....:|||||.....:
278  TCCAGGTCGACTTAAAGCTGTGCAACGCTTTGACGCT..... 315
105  GlnLysLysLeuProGluArgLysThrAlaIleValThrSerGlySerP 121
    |||||.....:|||||.....:
316  .....CTACCAAAAGAGAAATGGCGTCGGCACTCCGGTACCGG 356
121  oTyrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysP 138
    .....:|||||.....:
357  TGATATGGCACAAAATGGTTCGACATCTG.....GGAATCAGAGAC 400
138  rGlyValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
    |||||.....:|||||.....:
401  CAAAGTACTTCATTACCGCTAATGATGTCAAACAGGTAAGCTCATCA 450
155  GluGlyTyrSerArgAlaArgAspLeuLeuArgGlnAspLeuGlnLeu 171
    ||| .....:|||||.....:
451  GAACCATATCTGAAGCGCAGAGATGGCTTAGATATCCGATCAATCAGCA 500
171  rGlyLysGlnAspLeuLysTyrValValPheGluAspAlaProValGly 188
    .....:|||||.....:
501  AGACCTTCTCAAAATCTAAGGTAGTATTGAAGACGCTCCAGCAGTA 550
188  leLysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
    || .....:|||||.....:
551  TTGCCGCGCGAAAGCCGCGGTGTAAGATCATTTGATTCACACTACT 600
205  TyrAspLysSerValLeuPheAspAlaGlyAlaAspTyrValValCysAs 221
601  TTCGACTTGGACTTCTCTAAGAGAAAAGCC.....TGCA 635
221  pleuThrGlnValSerValValLysAsnAsnGlnAsn 233
    ||| .....:|||||.....:
636  CATC.....ATTGTCAAAAACCGACGAAATCC 660

seq_name: /SIS1/gcgdata/geneseq/geneseqn/NA1999.DAT: AAX12984

seq_documentation_block:
ID   AAX12984 standard; DNA: 17087 BP.
XX
AC   AAX12984.
XX
DT   19-MAR-1999 (first entry)
XX
DE   Enterococcus faecalis genome contig SEQ ID NO:47.
XX
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
XX
OS   Enterococcus faecalis.
XX

```

```

PN  M09850555-A2.
XX
PD  12-NOV-1998.
XX
XX  04-MAY-1998; 98WO-US08985.
XX
PR  14-NOV-1997; 97US-0066009.
PR  06-MAY-1997; 97US-0044031.
PR  16-MAY-1997; 97US-0046655.
XX
PA  (HUMA-) HUMAN GENOME SCI INC.
PI  Barash SC, Dillon PJ, Kunsch CA;
XX  WPI, 1999-045171/04.
DR
XX  New isolated Enterococcus faecalis polynucleotides and polypeptides
PT  - used to develop products for the detection of Enterococcus and for
PT  use in vaccines for prevention or attenuation of Enterococcus
PT  infection.
XX
PS  Claim 1; Page 429-438; 2084pp; English.
XX
CC  A computer readable medium has been developed which has recorded on it
CC  982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC  AAX12938 to AAX13919 represent these nucleotide sequences which are
CC  primary nucleotide sequences, also known as contigs. The computer-based
CC  system can identify fragments of the Enterococcus faecalis genome with
CC  commercial importance. The products can be used to detect the presence
CC  of Enterococcus faecalis in samples. They can also be used for
CC  diagnosing Enterococcal infection in an animal and monitoring
CC  progression of disease, and for identifying agents which can be used to
CC  modulate the growth or pathogenicity of Enterococcus faecalis, or
CC  another related organism, in vivo or in vitro. In particular the
CC  polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC  can be used in vaccines to prevent or attenuate an Enterococcal
CC  infection.
XX
SQ  Sequence 17087 BP; 4734 A; 3683 G; 3245 G; 5415 T; 10 other;

alignment_scores:
    Quality: 155.50      Length: 240
    Ratio: 1.341        Gaps: 9
    Percent Similarity: 48.333    Percent Identity: 27.083

alignment_block:
US-09-558-284-2 x AAX12984 ..

Align seg 1/1 to: AAX12984 from: 1 to: 17087

10  LeuPheAspLeuAspGlyThrIleValSerThrValAlaIaGluIu 26
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2773 ATTTTGATTGTGATGTGCTACTGACACGCCCAATTCATTATCA 2822

26  salatrPThrLysLeuCysTyrGluTyrGlyValAspProSerGluLeuP 43
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||
2823 AGCTTGAAAGCGTTGCGCATTCGCTAGTATCCCATGACAAACCT 2872

43  heLysHisSer.....
    ||:|||||:|||||:|||||:|||||:|||||:|||||:
2873 TCACACGAAACATTAAAGCATTAGTCGATGATTCCTTGACCGTATT 2922

47  .....HisGlyAlaArgThrGlnGluValIleuArgArgPheProIu 61
    ||||||:|||||:|||||:|||||:|||||:|||||:|||||
2923 TTAGTCCATGCGCACCGCGAA.....AATGCGTTTACTCCAGC 2960

61  sleuAspAspThrAspAsnLysGlyValLeuAlaLeuGluLysAspIleA 78
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||
2961 A.....GAAGAAAGACGCTTAGCTCAGCAAAAAAATGANC 2995

78  laHis.....SerTyrLeuAspThrValSerLeuIleProGly 90
    :||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

2996 ACTACGTTCAATTATTAGAACAAATTACACAGAGATGTTTACCTGGT 3045
91  AlagluAsnLeuLeuSerLeuAspValAspThrGluThrGlnLysIu 107
    :||:|||||:|||||:|||||:|||||:|||||:|||||
3046 GTCGTGCCATTGCTA.....CAGCAACGCCAAGCAGC 3077

107  sleuProGluArgLysTrpAlaIleValIhrSerGlySerProTyrLeuA 124
    :||:|||||:|||||:|||||:|||||:|||||:|||||
3078 TCACATTCCTCGCGCGCGTGGCTTCGCTCAAAAAACGCCGCC..... 3120

124  laPheSerTrpPheGluThrIleLeuLysAsnValGlyLysProLysVal 140
    ||||||:|||||:|||||:|||||:|||||:|||||:|||||
3121 .....CTGATTTTGAAAAAATTAGCGCGTCCGCGCTAC 3153

141  PheIleThrGlyPheAsp.....ValLysAsnGlyLysProAspPr 154
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3154 TTCGCCACCAATTGTGATGCCGACTGCTTAAGTAAAGCAAACTGATTC 3203

154  oGluGlyTyrSerArgAlaArgAspLeuLeuArgGlnAspLeuGlnLeuT 171
    ||||:|||||:|||||:|||||:|||||:|||||:|||||
3204 TGAATCTTTTATAGCCGCCGCTGACACTATT..... 3234

171  hrGlyLysGlnAspLeuLysTyrValAlaPheGluAspAlaProValGly 187
    ||||:|||||:|||||:|||||:|||||:|||||:|||||
3235 ..GCCGTGTAACGCCAANAAGCCCATTTGCTTTGAGATGACAAATCAGC 3282

188  IleLysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSe 204
    ||||:|||||:|||||:|||||:|||||:|||||:|||||
3283 ATTGACGCGTTGAAGCGCGCGCATCTATGCCGCTTGCGGCGCAG 3332

204  rTyrAspLysSerValLeuPheAspAlaGlyAlaAspTyrValValCysA 221
    |||||:|||||:|||||:|||||:|||||:|||||:|||||
3333 CCAACCTTA.....ATCGAGCGCGACATGCACAACTTCTG 3367

221  spleuThrGlnValSerVal 227
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||
3368 AATGACTGAACCTAGCGCTT 3387

seq_name: /SIDSI/gcdata/geneseq/geneseq/NA1996.DAT:AA142063

seq_documentation_block:
ID  AAT42063 standard; DNA; 1830121 BP.
XX
AC  AAT42063;
XX
DT  14-SEP-1999 (first entry)
XX
DE  Haemophilus influenzae complete genome sequence.
XX
KW  Genome; bacterium; Haemophilus influenzae; computer readable medium;
KW  expression modulating fragment; regulation; gene expression; vector;
KW  organism; open reading frame; ORF; ds.
XX
OS  Haemophilus influenzae.
XX
PN  W09633276-A1.
XX
PD  24-OCT-1996.
XX
PF  22-APR-1996; 96WO-US05320.
XX
PR  07-JUN-1995; 95US-0487429.
PR  21-APR-1995; 95US-0426787.
PR  07-JUN-1995; 95US-0476102.
XX
PA  (HUMA-) HUMAN GENOME SCI INC.
PA  (UTOO ) UNTV JOHNS HOPKINS.
XX
PI  Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
XX  WPI, 1996-485782/48.
DR
XX  Haemophilus influenzae Rd genome recorded on computer readable
PT

```

PT medium - useful for identifying commercially important nucleic acid
PT fragments by homology searching
PS Claim 1: Page 77.2-77.1091: 1291pp: English.
XX
CC This sequence represents the complete genome sequence of the bacterium
CC Haemophilus influenzae strain Rd. The invention relates to a computer
CC readable medium (CRM) having recorded upon it the complete H. influenzae
CC nucleotide sequence (1), a representative fragment of (1) or a nucleotide
CC sequence at least 99% identical to (1). By providing the full-length
CC genomic sequence in a computer readable form, it is possible to identify
CC commercially important nucleic acid fragments and expression modulating
CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
CC regulate the expression of a nucleic acid molecule. Vectors and altered
CC organisms comprising the predicted ORFs can be used to produce any of the
CC polypeptide fragments of the H. influenzae Rd genome.
SQ Sequence 1830121 BP: 567399 A: 350615 C: 347389 G: 564036 T: 682 other:

alignment_scores:
Quality: 151.00 Length: 206
Ratio: 1.411 Gaps: 7
Percent Similarity: 51.942 Percent Identity: 23.301

alignment_block:
US-09-558-284-2 x AAT42063 ..

Align seg 1/1 to: AAT42063 from: 1 to: 1830121

10 LeupheaspLeuaspGlyThrIleValSerThrValAlaIleAluLy 26
507488 ATTTTGTATGATGATGAACTTAAATGATACATGCCCGCATGCACA 507537
26 SAIArThrIleuLeuGlyTyrGlyValAspProSerGluLeuP 43
507538 AGCTTGACATGCTGGCAAGAAATTTGGCTATGAAATTTTTCAAA 507587
43 heLysHisSerHisGlyAlaIleThr.....GlnGluVal 54
507588 TTATGCTATTAATCTCGCGCGGCGACGTTCAACTATCGCGGGAATG 507637
55 LeuArgArGhePheProLysLeuAspThrAspAsnLysGlyValle 71
507638 ATGAACGCGCGCATATGCCACTAGATGATGATGAGAT.....GTGCT 507681
71 uAlaLeuGlnLysAspIleAlaHisSerTyrLeuAspThrValSerLeuI 88
507682 GCGTCGCAAAAGTGAGTTGCTTAT.....CAACTTA 507713
88 leProGlyAlaGlnAsnLeuLeuSerLeuAspValAspThrCluThr 104
507714 TTCCGACTCAATCAAGTATTTGCCACACTTTGAAAT..... 507751
105 GlnLysLysLeuProGluArgLysTrpAlaIleValThrSerGlySer.. 120
507752 GTTAATTCCTTCATCAGAAAAACCGCATCTTAGCTCTGCTGCACA 507801
121ProTyrL 123
507802 CCCTAAATTAATGATATGTTGATGATGAAGTGGCTATCGACCTTACT 507851
123 euAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysProLys 139
507852 TT.....AAT 507856
140 ValPheIleThrGlyPheAspValLysAsnGlyLysProAspProGluG 156
507857 GCAATTCGACACGATGATGATAAAGAACATAAACACACACAGAAAC 507906
156 yTyrSerArgLalaArgAspLeuLeuArgGlnAspLeuInleuThrGlyL 173
507907 TTTTTCAGTTGTGCGGAGTTAATTCAGCAAAAT..... 507940

173 ySGlnAspLeuLysTyrValValPheGlnAspAlaProValGlyLys 189
507941CCAAGTCGTGTATCGTATTTGAAGATGCGGATTTAGCGCTGCA 507985
190 AlagilyLysAlaMerGly 195
507986 GCAGGATTTAAGTGCAGGA 508003

seq_name: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:AA13135

seq_documentation_block:
ID AAX13135 standard; DNA: 1521 BP.
XX
AC AAX13135:
XX
DT 19-MAR-1999 (first entry)
XX
DE Enterococcus faecalis genome contig SEQ ID NO:198.
XX
KW Enterococcus faecalis; contig: detection; Enterococcal infection;
KM vaccine; attenuation; computer readable medium; ds.
XX
OS Enterococcus faecalis.
XX
PN W09850555-A2.
XX
PD 12-NOV-1998.
XX
PE 04-MAY-1998; 98WO-0508985.
XX
PR 14-NOV-1997; 97US-0066009.
PR 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046655.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Dillon PJ, Kunsch CA;
XX
DR WPI; 1999-045171/04.
XX
XX
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX
PS Claim 1; Page 1051-1052: 2084pp: English.
XX
XX A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAX12938 to AAX13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX
SQ Sequence 1521 BP: 458 A: 322 C: 258 G: 480 T: 3 other:

alignment_scores:
Quality: 143.00 Length: 250
Ratio: 1.059 Gaps: 11
Percent Similarity: 54.000 Percent Identity: 26.400

alignment_block:
US-09-558-284-2 x AAX13135/rev ..

Align seg 1/1 to reverse of: AAX13135 from: 1 to: 1521

```
10 LeuheaSpLeuaspGlyThrIleValSerThrThrValAlaIaGlyLeu 26
   :::::::::::::::::::::
1497 ATTTTGAATGATGATGGTGGTCTTTTGTATGACGAGTTCATTATAC 1448
26 salatRPhrLysLeuCysTyrGluTyrGlyValAspProSer...GluL 42
   :::::::::::::::::::::
1447 ATCTACCAAAAGTAGCATGATGGGTTTCCATATAGTAAAGAG 1398
42 eupheLysHisSerHisGlyAlaArgThrGluVal.....Leu 55
   :::::::::::::::::::::
1397 TCATATTAGACTATGTTGGATTTCGATGACAGAGTACAGAGAAATTAC 1348
56 ArgArpPhePheProLysLeuAspAspThrAspAsnLysGlyValLeuAl 72
   :::::::::::::::::::::
1347 CGACGATTATATGCA.....TCCTATGCTCATGACAC 1316
72 aLeuGluLysAspIleAlaHisSerTyrLeuAspThr..... 84
   :::::::::::::::::::::
1315 AGTGGAGAGAAATTATCCGCGTCTTATGATGATACCTGCAGAGTTTC 1266
85 .....ValSerLeuIleProGlyAlaGluAsnLeuLeuSer 97
   :::::::::::::::::::::
1265 GATCTGTAAAGTCCCTTAAACACAGGGGTGTGGAATTCCTGACTTT 1216
98 LeuAspValAspThrGluThrGluLysLeuProGluArgLysTrpAl 114
   :::::::::::::::::::::
1215 TTGAT.....GATCAAAAATTCCTGCA..... 1192
114 aIleValThrSerGlySerProTyrLeuAlaPheSerTrpPheGluThrI 131
   :::::::::::::::::::::
1191 .TTAGTGTCTCAAGTACGCTCGTCACGA.....ATTGAATGT 1152
131 leuLysAsnValGlyLysProLysValPhe.....IleThrGly 144
   :::::::::::::::::::::
1151 TATTAAACCCTGCTGATTCAGAGATGTTTGTAGTATGTTTCAGCA 1102
145 PheAspValLysAsnGlyLysProAspProGluGlyTyrSerArgAlaAr 161
   :::::::::::::::::::::
1101 GAGGATGTAAACGACGAAACCGGACCCGAAATTTTCAAAAGCTCG 1052
161 gaSpLeuLeuArgGlnAspLeuGlnLeuThrGlyLysGlnAspLeuLysT 178
   :::::::::::::::::::::
1051 T.....CACTTTTACGACAGACCCGCAAAA 1023
178 yValValPheGluAspAlaProValGlyIleLysAlaGlyLysAlaMet 194
   :::::::::::::::::::::
1022 CCTTGATTTTCGAGACTCATTTTCATGGGGTAAAGCGGCTCATAGTCG 973
195 GlyAlaIleThrValGlyIleThrSerSerTyrAspLysSerValLeuPh 211
   :::::::::::::::::::::
972 GGCATTCCAGTGCATGATGATGATTTATTCACACCACTCAATATAT 923
211 eaSpAlaGlyAlaAspTyrValValLysAspLeuThrGlnVal.SerVal 227
   :::::::::::::::::::::
922 CCAAGGAAAACATTACATGCTCTTGAGAGTTTACATCAAGCCCATCAT 873
228 ValLysAsnAsnGluAsnGlyIleValIleGlnValAsnAspProLeu 243
   :::::::::::::::::::::
872 ATTTAAATATATATAAAC.....AACACTCCTTTG 843
seq_name: /SIDS1/gcgdata/geneseq/NA2000.DAT: AAC45665
seq_documentation_block:
ID AAC45665 standard; DNA: 1057 BP.
XX
AC AAC45665:
XX
XX 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 47317.
```

```
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW Protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX EP1033405-A2.
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139482.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 18-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
```


PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0155658.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.

PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161820.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:

Quality: 136.00 Length: 221
Ratio: 1.183 Gaps: 7
Percent Similarity: 52.036 Percent Identity: 25.792

alignment_block:

US-09-558-284-2 x AAC38884 ..

Align seg 1/1 to: AAC38884 from: 1 to: 1060

10 LeupheAspleuAspGlyThrIleValSerThrThrValAlaIleAlaGly 26
|||||
182 CTGTTGATGTTGATGACACACCTCTGTCACATCCACATCCACCTAT 231
26 sAlArpThrIleuScySTyGluTyGly.....V 37
|||||
232 TGCCTTCCAGAACCTCTCAAGATGTTTAAACATGGGCTCCAA 281
37 aLasproSerCluBheIyHisSerHisGlyAlaArgThrGlu 53
|||||
282 TCGATGACAAATCTTGTGTGACAACTTGTCTGGAACACACATTCGAA 331
54 ValIeuArGArGpHePheProIyLeuAspAspThrAspAsnIyGly 70
|||||
332 ATTGCTGACTTCTGTCTCT.....GATGATGTTTCAAGAGGTTTAA 375
70 lleuAlaIeuGluIyAspIleAlaHisSerTyIleuAspThrValSer 87
|||||
376 ATTCTGTGATGAAAGAACCTCTTACCCGAAATTTCTTCAGAGAAGA 425
87 euIleProGluAlaGluAsnIeuIeuSerIeAspValAspThrGlu 103
|||||
426 TAAAGCCA.....CTTGATGGCTTATATAA 451
104 ThrIlnIySleuProGluArg.....IySTrPaIaIleAlaThrSe 118
|||||
452 CTGACCAAAATGATCGAAGATCGTGAATGAACGATCGCGTAACAA 501
118 rGlySerProTyIleuAlaPheSerTrpPheGluThrIleuIySasv 135
|||||
502 CGCTCCAAAGAAAGCA.....GAGCTCATGATATCGAAGAAC 539
135 aIcIyIySProIyValPhe.....IleThrGlyPheAspValIyS 148
|||||
540 TTGGTCTGACTGATTTCTTTCAGACAGATGTTTGGCTGTGATGGA 589
149 AsnGlyIySProAspProGluGlyTyIleSerArgAlaIleArgAspIleuAr 165

```

590 TTTCCAAACACACCCCTGACCTTACTTGAAGCCTCTGAAGTCTTA 639
      |||||  |||  |||  ::|||  ::::|||:
165 gGlnAspLeuGlnLeuThrGlyLysGlnAspLeuLysTyrValValPheG 182
      ::::  ::::  |||  ::||||||
640 CGTGTCAAGGAGCACACG.....CTAGTTTTCG 668
      ::::  ::::  |||  ::||||||
182 LAspAlaProValGlyIleLysAlaGlyLysAlaMetGlyAlaIleThr 198
      ||||:  ::  |||||  |||||  |||  |||  ::
669 AAGACTCGATCTCTGGATAAAGCTGAGTTGCAGCTGSAATGCCAGTG 718
      ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
199 ValGlyIleThrSerSerTyrAspLysSerValLeuPheAspAlaGlyAl 215
      ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
719 ATGGAGCTGACTACAGGAATCCAGCAAGCCTTATGCAACAAAC 768
      ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
215 aAspTyrValVal 219
      ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
769 CGCTTTCTCATC 781

```


OM of: US-09-558-284-2 to: Issued_Patents_NA:* out_format : pfs
Date: Jul 22, 2001 9:28 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US09558284/rtnat.20072001.160145_19160/app-query.fasta.1.306
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2n.rn1
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPECL=0.000
-LOOPEXT=0.000 -QCAPOP=4.500 -QCAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YCAPOP=10.000
-YGAEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45 -DOCALLIGN=200
-THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09558284@cgn1_1.60 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1

Search information block:

Query: US-09-558-284-2
Query length: 246
Database: Issued_Patents_NA:*
Database sequences: 317530
Database length: 92630169
Search time (sec): 62.610000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation	
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-08-968-563-8 +		328.50	751.09	3.1e-34	816	1
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-08-969-683A-8 +		328.50	751.09	3.1e-34	816	1
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-08-968-563-9 +		327.50	749.69	3.8e-34	753	1
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-08-969-683A-9 +		327.50	749.69	3.8e-34	753	1
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-08-998-416-874 +		118.00	253.74	1.6e-06	590	1
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-08-998-416-874 +		109.50	223.11	8.0e-05	1380	1
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-463-288A-14 +		109.50	223.11	8.0e-05	1380	1
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-470-445A-14 +		109.50	223.11	8.0e-05	1380	1
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-462-679-14 +		109.50	223.11	8.0e-05	1380	1
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-466-210A-14 +		109.50	223.11	8.0e-05	1380	1
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-467-147A-14 +		109.50	223.11	8.0e-05	1380	1
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-469-014-14 +		109.50	223.11	8.0e-05	1380	1
/cgn2_6/prodata/2/ina/PCTUS.COMB.seq:PCT-US93-10506A-14 +		109.50	223.11	8.0e-05	1380	1
/cgn2_6/prodata/2/ina/PCTUS.COMB.seq:PCT-US93-10506-14 +		109.50	223.11	8.0e-05	1380	1
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-937-102-1 +		91.00	167.19	0.1047	3639	1
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-876-991-1 +		91.00	152.31	0.7059	12284	1
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-09-059-853-1 +		91.00	152.31	0.7059	12284	1
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-08-750-717-1 +		91.00	152.29	0.7053	12311	1
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-07-961-522-3 +		86.50	164.40	0.1497	1903	1
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-217-438-3 +		86.50	164.40	0.1497	1903	1
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-321-978-3 +		86.50	164.40	0.1497	1903	1
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-710-584-3 +		86.50	164.40	0.1497	1903	1
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-08-904-263A-3 +		82.50	146.34	1.52	3825	1
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-119-125A-3 +		82.50	145.44	1.71	4118	1
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-09-433-262-1 +		82.50	130.40	11.73	14078	1
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-08-859-694-1 +		82.50	129.98	12.33	14578	1
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-892-770-2 +		80.50	149.95	0.9559	1929	1
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-09-162-372-2 +		79.50	138.86	3.96	3930	1
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-743-637B-27 -		79.50	128.59	14.79	9100	1
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-08-526-840B-27 -		79.50	128.59	14.79	9100	1
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-08-501-126-1 +		78.00	145.89	0.9629	1191	1
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-09-132-028-1 +		78.00	142.92	2.35	2106	1
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-477-451-9 +		77.50	129.77	12.71	5599	1
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-477-451-13 -		77.50	129.77	12.71	5599	1
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-323-170B-1 +		76.50	120.75	40.44	9636	1
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-08-998-416-108 +		75.50	152.14	0.7216	609	1
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-279-754-1 +		75.50	147.12	1.37	918	1
/cgn2_6/prodata/2/ina/PCTUS.COMB.seq:PCT-US95-09052-1 +		75.50	147.12	1.37	918	1
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-458-067-1 +		75.50	145.48	1.70	1050	1
/cgn2_6/prodata/2/ina/PCTUS.COMB.seq:PCT-US96-07795-1 +		75.50	145.48	1.70	1050	1

/cgn2_6/prodata/2/ina/PCTUS.COMB.seq:PCT-US96-07796-1 + 75.50 145.48 1.70 10
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-07-832-855-1 + 75.50 138.19 4.32 1905
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-760-335A-1 + 75.50 137.91 4.47 1949
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-09-318-448-27 + 75.00 134.00 7.40 2436
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-095-737-3 + 75.00 131.31 10.43 3033

seq_name: /cgn2_6/prodata/2/ina/6A.COMB.seq:US-08-968-563-8

seq_documentation block:

Sequence 8, Application US/08968563
Patent No. 6013494
GENERAL INFORMATION:
APPLICANT: CHARLES E. MAKAMURA
APPLICANT: ANTHONY A. GATENBY
APPLICANT: AMY (KUANG-HUA) HSU
APPLICANT: RICHARD D. LA REAU
APPLICANT: SHARON L. HAYNIE
APPLICANT: MARIA DIAZ-TORRES
APPLICANT: DONALD E. TRIMBUR
APPLICANT: GREGORY M. WHITED
APPLICANT: VASANTHA NAGARAJAN
APPLICANT: MARK S. PAYNE
APPLICANT: STEPHEN K. PICATAGGIO
APPLICANT: RAMESCH V. NAIR
TITLE OF INVENTION: METHOD FOR THE RECOMBINANT
TITLE OF INVENTION: PRODUCTION OF 1,3-PROPANEDIOL
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: U.S.A.
ZIP: 19898
ADDRESSEE: GENENCOR INTERNATIONAL, INC.
STREET: 4 CAMBRIDGE PLACE
STREET: 1870 SOUTH WINTON ROAD
CITY: ROCHESTER
STATE: NEW YORK
COUNTRY: U.S.A.
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.50 INCH DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
SOFTWARE: MICROSOFT WORD VERSION 7.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968,563
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/030,601
FILING DATE: NOVEMBER 13, 1996
ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA AXAMETHY
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CR-9982
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 816 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Gp1
US-08-968-563-8

alignment_scores: 328.50 Length: 229

Ratio: 2.313 Gaps: 8
Percent Similarity: 62.009 Percent Identity: 37.555
alignment_block:
US-09-558-284-2 x US-08-968-563-8 ..

Align seg 1/1 to: US-08-968-563-8 from: 1 to: 816

```

10 LeuPheAspLeuSpGlyThrIleValSerThrValAlaIaGluVal 26
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109 CTATTGATGATTGACGCTACCATCATCATCTCTCAACACCATGCTGC 158
26 salATrPThrLysLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 38
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
159 TTCTCTGAGA.....GATTTCGCTAAAGACACCTTACTTCG 196
38 sPProSeGluLeuPheLysHisSerHisGlyAlaArgThrGlnGluVal 54
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
197 ATGCCGACACGTTATTACATCTCTCAGCGTTGAGACACTTACGATGCC 246
55 LeuArgArgPhePheProLysLeuAspPThrAspAsnLysGlyValle 71
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
247 ATTGCCAAGTTCGCTCCA.....GACTTTCGTCGATCAGACATACGTTAA 290
71 uAlaLeuGluLysAspIleAlaHisSerTyrlLeuAspPThrValSerLeu 88
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
291 CAAGCTAGAAGGTGAATCCAGAAAGTACGGGTACGACTCCATCGTAG 340
88 lProGlyAlaGluAsnLeuLeuSerLeuAspValAspPThrGlnThr 104
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
341 TTCAGAGTGTCTCAAGTGTGTAACTGTTGAACGCC..... 378
105 GlnLysLysLeuProGluArgLysTrpAlaIleValThrSerGlySerP 121
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
379 .....TTGCCAAAGGAAAAATGGCGTGTGCCACCTCGTACCG 419
121 oTyrlLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysP 138
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
420 TGCATGCGCCAGAAATGTTGCAC...ATTTCGAG...ATCAGAGAC 463
138 rOlyValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
464 CAGAACTACTTCATCACCAGCAATGATGTCAAGCAAGTAAAGCTCACCCA 513
155 GluGlyTyrlSerArgAlaArgAspLeuLeuArgGlnAspLeuLeuThr 171
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
514 GAACCATACTTAAGGAGTAAACGCTTGGCTTCCCATTTATGACACA 563
171 rGlyLysGlnAspLeuLysTyrlValValPheGluAspAlaProValGly 188
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
564 AGACCCATCCAAATCTAAGGTGTGCTTTGAAGACGCGACGCTGTGTA 613
188 lElyAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
614 TTCTCTCTGTTAAGCTGCTGCTGTTAAATCGTGTATGCTACACAT 663
205 TyrlAspLysSerValLeuPheAspAlaGlyAlaAspTyrlValLysAs 221
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
664 TTCGATTTGGACTTCTTGAAGAAAGGCT.....TGTGA 698
221 pLeuThrGlnValSerValValLysAsnGlnGluAsn 233
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
699 CACT.....ATTGCTCAAGAACACGACGATCT 723
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-969-683A-8
seq_documentation_block:
; Sequence 8, Application US/08969683A
; Patent No. 6136576
; GENERAL INFORMATION:
; APPLICANT: GENENCOR INTERNATIONAL, INC.
; TITLE OF INVENTION: METHOD FOR THE RECOMBINANT
; TITLE OF INVENTION: PRODUCTION OF 1,3 PROPANEDIOL

```

```

? NUMBER OF SEQUENCES: 68
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Genencor International, Inc.
? STREET: 4 Cambridge place
? STREET: 1870 South Winton road
? CITY: Rochester
? STATE: NY
? COUNTRY: U.S.A
? ZIP: 14618
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: Windows
? SOFTWARE: FASTSEQ for Windows Version 2.0b
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/969,683A
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: PCT/US97/20873
? FILING DATE: 13-NOV-1997
? APPLICATION NUMBER: 60/030,601
? FILING DATE: 13-NOV-1996
? ATTORNEY/AGENT INFORMATION:
? NAME: Glaister, Debra
? REGISTRATION NUMBER: 33,888
? REFERENCE/DOCKET NUMBER: GC 369-2
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 650-864-7620
? TELEFAX: 650-845-6504
? TELEX:
? INFORMATION FOR SEQ ID NO: 8:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 816 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? ORIGINAL SOURCE:
? ORGANISM: GPPI
? US-08-969-683A-8

```

alignment_scores:
Quality: 328.50 Length: 229
Ratio: 2.313 Gaps: 8
Percent Similarity: 62.009 Percent Identity: 37.555

alignment_block:
US-09-558-284-2 x US-08-969-683A-8 ..

Align seg 1/1 to: US-08-969-683A-8 from: 1 to: 816

```

10 LeuPheAspLeuSpGlyThrIleValSerThrValAlaIaGluVal 26
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109 CTATTGATGATTGACGCTACCATCATCATCTCTCAACACCATGCTGC 158
26 salATrPThrLysLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 38
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
159 TTCTCTGAGA.....GATTTCGCTAAAGACACCTTACTTCG 196
38 sPProSeGluLeuPheLysHisSerHisGlyAlaArgThrGlnGluVal 54
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
197 ATGCCGACACGTTATTACATCTCTCAGCGTTGAGACACTTACGATGCC 246
55 LeuArgArgPhePheProLysLeuAspPThrAspAsnLysGlyValle 71
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
247 ATTGCCAAGTTCGCTCCA.....GACTTTCGTCGATCAGACATACGTTAA 290
71 uAlaLeuGluLysAspIleAlaHisSerTyrlLeuAspPThrValSerLeu 88
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
291 CAAGCTAGAAGGTGAATCCAGAAAGTACGGGTGAACACTCATCGTAG 340

```

```

88  leProGlyAlaGluAsnLeuLeuSerLeuAspValAspThr 104
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
341  TCCAGGCTGTCAGTGTGTAATCTTGAAGCC..... 378
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
105  GluLysLeuProGluArgLysTrpAlaIleThrSerGlySer 121
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
379  .....TTGCCAAAGAAATAATGGGCTGTGCCACCTCTGGTACCG 419
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
121  oTyrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysP 138
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
420  TGACATGCCAAGAATGTTCCGAC...ATTTCGAG...ATCAAGACAC 463
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
138  roLysValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
464  CAGAACTACTATCACCAGCATGATGTCAAGCAAGTAACCTCACCCCA 513
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
155  GluGlyTyrSerArgAlaArgAspLeuLeuArgGlnAspLeuGlnLeuTh 171
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
514  GAACCACTACTTAAAGGGTAGAAGGTTTGGTTCCCAATTATGAACA 563
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
171  rGlyGlnAspLeuLysTyrValValPheGluAspAlaProValGlyI 188
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
564  AGACCCATCCAAATCTAAGCTGTCTCTTGAAGACGCACGACGCTGTA 613
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
188  leLysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSer 204
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
614  TTGCTGCTGGTAAGGCTGTGGCTGTAATAATCGTTGATGTACACACT 663
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
205  TyrAspLysSerValLeuPheAspAlaGlyAlaAspTyrValGlySas 221
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
664  TTCGATTGGACTTCTTGAAGGAAAGGCT.....TGTGA 698
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
221  PleuThrGlnValSerValValLysAsnAsnGluAsn 233
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
699  CATC.....ATTGTCAGAAACACGACGATCT 723
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

seq_name: /cgn2.6/ptodata/2/lna/6A.COMB.seq:US-08-968-563-9

seq_documentation_block:

: Sequence 9, Application US/08968563

: Patent No. 6013494

: GENERAL INFORMATION:

: APPLICANT: CHARLES E. NAKAMURA

: APPLICANT: ANTHONY A. GATENBY

: APPLICANT: AMY (KUANG-HUA) HSU

: APPLICANT: RICHARD D. LA REAU

: APPLICANT: SHARON L. HAYNIE

: APPLICANT: MARIA DIAZ-TORRES

: APPLICANT: DONALD E. TRIMBUR

: APPLICANT: GREGORY M. WHITED

: APPLICANT: VASANTHA NAGARAJAN

: APPLICANT: MARK S. PAYNE

: APPLICANT: STEPHEN K. PICATAGGIO

: APPLICANT: RAMESH V. NAIR

: TITLE OF INVENTION: METHOD FOR THE RECOMBINANT

: NUMBER OF SEQUENCES: 49

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY

: STREET: 1007 MARKET STREET

: CITY: WILMINGTON

: STATE: DELAWARE

: COUNTRY: U.S.A.

: ZIP: 19898

: ADDRESSEE: GENENCOR INTERNATIONAL, INC.

: STREET: 4 CAMBRIDGE PLACE

: CITY: ROCHESTER

: STATE: NEW YORK

: COUNTRY: U.S.A.

: COMPUTER READABLE FORM:

: MEDIUM TYPE: 3.50 INCH DISKETTE

: COMPUTER: IBM PC COMPATIBLE

```

: OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
: SOFTWARE: MICROSOFT WORD VERSION 7.0A
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/968,563
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/030,601
: FILING DATE: NOVEMBER 13, 1996
: ATTORNEY/AGENT INFORMATION:
: NAME: FLOYD, LINDA AXAMETHY
: REGISTRATION NUMBER: 33,692
: REFERENCE/DOCKET NUMBER: CR-9982
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 302-892-8112
: TELEFAX: 302-773-0164
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 753 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORGANISM: GPP2
: US-08-968-563-9

```

```

alignment_scores:
  Quality: 327.50      Length: 229
  Ratio: 2.323        Gaps: 7
  Percent Similarity: 61.572  Percent Identity: 36.245

```

alignment_block:
US-09-558-284-2 x US-08-968-563-9 ..

Align seg 1/1 to: US-08-968-563-9 from: 1 to: 753

```

10  LeuPheAspLeuAspGlyThrIleValSerThrTrpValAlaIleGlu 26
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
46  TTGTTGACGCTGACGAGTACATTAATCTCTCAACACGACCATTCGTC 95
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
26  sAlaTrpThrLysLeuGlyTyrGluTyrGly.....ValA 38
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
96  ATTTCGAGG.....GATTTCGTAAGACAAACCTTATTTCG 133
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
38  sProSerGluLeuPheLysHisSerHisGlyAlaArgThrGlnGluVal 54
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
134  ATGCTGAACACGTTATCCAGCTCTGCAATGCTGGAGAAACGTTGATGCC 183
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
55  LeuArgArgPhePheProLysLeuAspThrAspAsnLysGlyValle 71
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
184  ATTGTAAGTTCGCTCCA.....GACTTTCGCAATGAAGATATGTTAA 227
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
71  uAlaLeuGlnLysAspIleAlaHisSerTyrLeuAspThrValSerLeu 88
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
228  CAAATTGAAGACTGAATTCCGGTCAAGTACGGTGAATAAATTCATTGAAG 277
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
88  leProGlyAlaGluAsnLeuLeuSerLeuAspValAspThrGluThr 104
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
278  TCCAGGCTGAGTTAAGCTGTGCAACGCTTGAACGT..... 315
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
105  GluLysLeuProGluArgLysTrpAlaIleValThrSerGlySerPr 121
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
316  .....TACCAAAAGACAAATGGGCTGTGCAACTTCCGCTACCGC 356
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
121  oTyrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysP 138
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
357  TGATATGCGCAAAATAATGCTTCGACATCTG.....GGAATCAGAGAC 400
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
138  roLysValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
401  CAAAGTACTTACTTAACGCTAATGATGTCAACAGGTAACGCTCATCA 450
   :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```



```

? APPLICANT: Pohlmann, Rainer
? APPLICANT: Steiner, Sabine
? APPLICANT: Mohr, Christine
? APPLICANT: Wendland, Jurgen
? APPLICANT: Knechtle, Philipp
? APPLICANT: Reischung, Corinne
? TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBVA GOSSTYPLI
? NUMBER OF SEQUENCES: 1152
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: No. 6239264rtis Corporation
? STREET: 3054 Cornwallis Road
? CITY: Research Triangle Park
? STATE: NO. 6239264th Carolina
? COUNTRY: USA
? ZIP: 27709
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/998,416
? FILING DATE: 24-DEC-1997
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: CH 0016/97
? FILING DATE: 31-DEC-1996
? ATTORNEY/AGENT INFORMATION:
? NAME: Meigs, J. Timothy
? REGISTRATION NUMBER: 38, 241
? REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 919-541-8587
? TELEFAX: 919-541-8689
? INFORMATION FOR SEQ. ID NO: 874:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 590 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? ORIGINAL SOURCE:
? ORGANISM: PAG15480P
? US-08-998-416-874

alignment_scores:
? Quality: 118.00 Length: 139
? Ratio: 1.494 Gaps: 6
? Percent Similarity: 56.835 Percent Identity: 30.216

alignment_block:
? US-09-558-284-2 x US-08-998-416-874 ..
?
? Align seg 1/1 to: US-08-998-416-874 from: 1 to: 590
?
? 10 leuPheaspLeuaspGLythrIleValSerThrhrValAlaIaIaGluVal 26
? |||||.....|||
? 222 CTATTCAGCTCGACGGACCATCATCTCGACGCCGCGTACAGGCC 271
?
? 26 salatrpThrLysLeuGlyTyrGluTyrGly.....ValA 38
? |||||.....
? 272 CTCTGTGAGG.....GAGTTGGCAAGACAGACCCCTACTTCCG 305
?
? 38 sPProSerGluLeuPheLysHisSerIscIyAlaArgThrGluGluVal 54
? ||.....
? 310 ATGCGGAGCATGTCATCATGCGCACCGCTGTGAGAACCTACAGCCC 359
?
? 55 leuAArgArgPhePheProLysLeuAspAspThrAspAsnLysGlyValIle 71
? :||: ||| ||| :|||
? 360 ATCGTACCTTCGCGGCA.....GACTATCTGAGTGGAGTACGTGAC 403
```

```

71 uAlaLeuGluLysAspIleAlaHisSerTyrLeuAspThrValSerLeuI 88
? |||||.....
? 404 GAGACTGTGAGGCGAATAATCCAGACAACTACGCCAAGTTCTCCGTGAGG 453
?
? 88 leProGlyAlaGluAsnLeuLeuLeuSerLeuAspValAspThrGluTh 104
? :|||||.....
? 454 TTCCCGGCGGTGACGCTGCAATTCCTTGAC..... 489
?
? 104 rGlnLysLysLeuProGluArgLysTrpAlaIleValThrSerGlySerP 121
? |||||.....
? 490 .....RACTTCGAGAGAAATGCGCGCTGCTACTTCGCGCCCT 532
?
? 121 roTyrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLys 137
? :|||.....
? 533 TCCAGATGCGACCC.AAGTGTTGATGTCCTC.....GCCATCAAGCGT 575
?
? 138 ProLysValPheIle 142
? |||||.....
? 576 CCTAGCACCTTCATT 590

seq_name: /cgn2_6/plodata/2/ina/5A_COMB.seq:US-08-363-311-14

seq_documentation_block:
? Sequence 14, Application US/08363311
? Patent No. 6548241
? GENERAL INFORMATION:
? APPLICANT: Michel, James L.
? APPLICANT: Ausubel, Frederick M.
? TITLE OF INVENTION: Conjugate Vaccine Against Group B
? NUMBER OF INVENTION: Streptococcus
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Sterne, Kessler, Goldstein & Fox
? STREET: 1225 Connecticut Avenue, N.W.
? CITY: Washington
? STATE: D.C.
? COUNTRY: USA
? ZIP: 20036-2678
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/363,311
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US/07/968,866
? FILING DATE: 02-NOV-1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Cimbalia, Michele A.
? REGISTRATION NUMBER: 33,851
? REFERENCE/DOCKET NUMBER: 0609.3740004
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 466-0800
? TELEFAX: (202) 833-8716
? INFORMATION FOR SEQ. ID NO: 14:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1380 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 79..1173
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 1004
? OTHER INFORMATION: /note="this feature is to signify
? OTHER INFORMATION: that the nucleotide sequence from position 757
? OTHER INFORMATION: through 1003 is inserted at position 1004 and
```

OTHER INFORMATION: may be repeated up to eight times (for a total of
 OTHER INFORMATION: nine repeating copies of these sequences within
 OTHER INFORMATION: the polynucleotide)."

alignment_scores:

Quality: 109.50 Length: 229
 Ratio: 0.928 Gaps: 11
 Percent Similarity: 51.528 Percent Identity: 24.891

alignment_block:

US-09-558-284-2 x US-08-363-311-14 ..

Align seg 1/1 to: US-08-363-311-14 from: 1 to: 1380

```

27 AlaTrpThrLysLeuCySerGlu...TyrGlyValAspProSerGlu 42
   |||:||||| ||| ||| |||
31 GCCTAATTTACTCTGTATAGAGTTATACAGATAAAGAGAAATATTAT 80
   |||:||||| ||| ||| |||
42 uphellyshisser.....HisGlyAlaArgThrGlnGluValLeuArg 57
   |||:||||| ||| ||| |||
81 GTTAGAAGAGCTTAAATAACAGTTATGATCTCACAGAGAAACAC 130
   |||:||||| ||| ||| |||
57 rgrhnerhrprrllyleuasprrhrasprhrrasprhrrasprhrr 73
   |||:||||| ||| ||| |||
131 GGTTCCTCAATTAG..... 144
   |||:||||| ||| ||| |||
74 GluLysAspIleAlaHisSerTyrLeuAspThrValSerLeuIlePro 90
   |||:||||| ||| ||| |||
145 .....AGTTCACAGTTGG 158
   |||:||||| ||| ||| |||
90 yAlaGluAsnLeuLeuSerLeuAsp...ValAspThrGlnThrGln 106
   |||:||||| ||| ||| |||
159 TGCACCTCTGTACTAATGCTTGTAGTTTGGTGGGTTACACAG 208
   |||:||||| ||| ||| |||
106 yLysLeuProGluArgLysThrAlaIleValThrSerGlySer...Pro 121
   |||:||||| ||| ||| |||
209 GTAACTTAATATTTTGAAGAGTAATAGTTGCGCATCTCAATTTCA 258
   |||:||||| ||| ||| |||
122 TyrLeuAlaPheSerTyrPheGluThrIleLeuLysAsnValGlyLys 138
   |||:||||| ||| ||| |||
259 GCGAGTCGACGACCTTAATACAGCATCAATAAATATACAAACCG 308
   |||:||||| ||| ||| |||
138 OlyValAlaPheIleThrGlyPheAspValLysAsnGlyLysProAsp 154
   |||:||||| ||| ||| |||
309 AAATGCTTACATAGATTATATGATGATTAATTAATTAATGATTCAT 358
   |||:||||| ||| ||| |||
155 .....GluGlyTyrSerArgAlaArgAspLeu 164
   |||:||||| ||| ||| |||
359 TACATTAATTTGTTTAGAACAAAGTTTACA...GCAGATATGTTT 405
   |||:||||| ||| ||| |||
165 ArgGln.....AspLeuGlnLeuThrGln 172
   |||:||||| ||| ||| |||
406 AGACAAGGACTAATAATAGTATGGGATGTTTCTCAGTTGACAGATGAC 455
   |||:||||| ||| ||| |||
172 yLysGlnAspLeuLysTyrValValPhe...GluAspAlaProValGly 188
   |||:||||| ||| ||| |||
456 AAGGCTAGTCTTACCTAATATATTTGGTGAAGATGACATACCCATG 505
   |||:||||| ||| ||| |||
188 LeuAlaAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSer 204
   |||:||||| ||| ||| |||
506 TAAAGACATGATGACAAATTTGATATAGTGTGCTTTAACT...ATT 552
   |||:||||| ||| ||| |||
205 TyrAspLysSerValLeuPheAspAlaGlyAlaAspTyrValValLys 221
   |||:||||| ||| ||| |||
553 TATGATTCACAACCTTGGAGGAT..... 576
   |||:||||| ||| ||| |||
221 pleuthrGlnValSerValValLysAsnGlnAsn 233
   |||:||||| ||| ||| |||
577 .....AAGATTGAAGAGTTAGAACGAATGCAAC 606
   |||:||||| ||| ||| |||

```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-463-288A-14

seq documentation block:

```

; Sequence 14, Application US/08463288A
; Patent No. 5820860
; GENERAL INFORMATION:
; APPLICANT: Michel, James L.
; APPLICANT: Kasper, Dennis L.
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Madoff, Lawrence C.
; TITLE OF INVENTION: Conjugate Vaccine For Group B
; TITLE OF INVENTION: Streptococcus
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, NW, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,288A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/363,311
; FILING DATE: 22-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/968,866
; FILING DATE: 02-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/408,036
; FILING DATE: 15-SEP-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Bugalsky, Lawrence B.
; REGISTRATION NUMBER: 35,086
; REFERENCE/DOCKET NUMBER: 0609,2370007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; TELEX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1380 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 79..1173
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 1004
; OTHER INFORMATION: /note="This feature is to signify
; that the nucleotide sequence from position 757
; through 1003 is inserted at position 1004 and can be
; OTHER INFORMATION: repeated up to eight times (for a total of nine
; OTHER INFORMATION: repeating copies of these sequences within the
; OTHER INFORMATION: polynucleotide)."
; US-08-463-288A-14

```

alignment_scores:

Quality: 109.50 Length: 229
 Ratio: 0.928 Gaps: 11
 Percent Similarity: 51.528 Percent Identity: 24.891

alignment_block:

US-09-558-284-2 x US-08-463-288A-14 ..

Align seq 1/1 to: US-08-463-288A-14 from: 1 to: 1380

```

27 AlatrphrlysluGcysTyrGlu...TyrGlyValAspProSerGluLe 42
   |||:||||| ||| ||| |||:
31 GCCATATATTACTCTGTATAGACTTATACAGACTAAAGGAGAAATTATTAT 80
   |||:||||| ||| ||| |||:
42 uphelyshisSer.....HisGlyAlaArgThrGlnGluValLeuArgA 57
   |||:||||| ||| ||| |||:
81 GTTAGAAGGCTAAATAATACACTTATGATCTTACACAGCAAGCAACAC 130
   |||:||||| ||| ||| |||:
57 rghphepProlysluAspAspThrAspAsnlySGlyValLeuAlaLeu 73
   |||:||||| ||| ||| |||:
131 GGTTCATTTAAAG..... 144
   |||:||||| ||| ||| |||:
74 GlulysAspLeuAlaHisSerTyrLeuAspThrValSerLeuIleProG 90
   |||:||||| ||| ||| |||:
145 .....AAGTTCAAGCTTGG 158
   |||:||||| ||| ||| |||:
90 yAlaGluAsnLeuLeuLeuSerLeuAsp...ValAspThrGlnThrGlnL 106
   |||:||||| ||| ||| |||:
159 TCGAGGCTTCTGTACTAATGGTCTTACTTTTGGTGGGCTTACACAG 208
   |||:||||| ||| ||| |||:
106 yslYsluLeuProGluArgLysTrpAlaIleValThrSerGlySer...Pro 121
   |||:||||| ||| ||| |||:
209 GTATATCTTAATATTGTTGAAAGAGCTCAATAGTCTGATCTACAAATTCCA 258
   |||:||||| ||| ||| |||:
122 TyrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysPr 138
   |||:||||| ||| ||| |||:
259 GCGAGCTCAGCAGCCTTAATACAGAGCATCAGTAAATATACAAACGCG 308
   |||:||||| ||| ||| |||:
138 olYsValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro. 154
   |||:||||| ||| ||| |||:
309 AAANGCTTACATGATTTATATGATGTAATAATTAGTAATAATAGATCCAT 358
   |||:||||| ||| ||| |||:
155 .....GluGlyTyrSerArgAlaArgAspLeuLeu 164
   |||:||||| ||| ||| |||:
359 TACAATTAATGTTGTTTAGAACACAGTTTACA...GCCAACTATGTTT 405
   |||:||||| ||| ||| |||:
165 ArgGln.....AspLeuGlnLeuThrGln 172
   |||:||||| ||| ||| |||:
406 AGACACAGCTACTAATCTATGGGAGCTTCTCTAGTTGCCAGACTACAGG 455
   |||:||||| ||| ||| |||:
172 yLysGlnAspLeuLysTyrValValPhe...GluAspAlaProValGlyI 188
   |||:||||| ||| ||| |||:
456 AAGGAGTACTTACTATATATATTTGGTGAAGATGACTTACCACATG 505
   |||:||||| ||| ||| |||:
188 lelysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
   |||:||||| ||| ||| |||:
506 TAAAGACTGATGCAACAATTGATATAGTCTTCTTAACT...ATT 552
   |||:||||| ||| ||| |||:
205 TyrAspLysSerValLeuPheAspAlaGlyAlaAspTyrValValLysAs 221
   |||:||||| ||| ||| |||:
553 TATGATTCAACACCTTGAGCGCAT..... 576
   |||:||||| ||| ||| |||:
221 PleuThrGlnValSerValValLysAsnAsnGluAsn 233
   |||:||||| ||| ||| |||:
577 .....AAGATTGAAGAGACTTAGAAGCAATGCCAAC 606
   |||:||||| ||| ||| |||:

```

seq_name: /cgn2.6/plodata/2/ina/5B_COMB.seq:US-08-470-445A-14

seq_documentation_block:

Sequence 14, Application US/0847045A

Patent No. 5833444

GENERAL INFORMATION:

APPLICANT: Michel, James L.

APPLICANT: Kasper, Dennis L.

APPLICANT: Ausubel, Frederick M.

APPLICANT: Madoff, Lawrence C.

TITLE OF INVENTION: Conjugate Vaccine Against Group B

NUMBER OF SEQUENCES: 65

CORRESPONDENCE ADDRESS:

```

ADDRESS: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, NW, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,445A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/363,311
FILING DATE: 22-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/968,866
FILING DATE: 02-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/408,036
FILING DATE: 15-SEP-1989
ATTORNEY/AGENT INFORMATION:
NAME: Bugalsky, Lawrence B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 0609.237000A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
TELEX: 248636 SSK
INFORMATION FOR SEO ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1380 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 79..1173
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1004
OTHER INFORMATION: /note="this feature is to signify
OTHER INFORMATION: that the nucleotide sequence from position 757
OTHER INFORMATION: through 1003 is inserted at position 1004 and can be
OTHER INFORMATION: repeated up to eight times (for a total of nine
OTHER INFORMATION: repeating copies of these sequences within the
OTHER INFORMATION: polynucleotide)."
```

US-08-470-445A-14

alignment_scores:

Quality: 109.50 Length: 229

Ratio: 0.928 Gaps: 11

Percent Similarity: 51.528 Percent Identity: 24.891

alignment_block:

US-09-558-284-2 x US-08-470-445A-14 ..

Align seq 1/1 to: US-08-470-445A-14 from: 1 to: 1380

```

27 AlatrphrlysluGcysTyrGlu...TyrGlyValAspProSerGluLe 42
   |||:||||| ||| ||| |||:
31 GCCATATATTACTCTGTATAGACTTATACAGACTAAAGGAGAAATTATTAT 80
   |||:||||| ||| ||| |||:
42 uphelyshisSer.....HisGlyAlaArgThrGlnGluValLeuArgA 57
   |||:||||| ||| ||| |||:
81 GTTAGAAGGCTAAATAATACACTTATGATCTTACACAGCAAGCAACAC 130
   |||:||||| ||| ||| |||:
57 rghphepProlysluAspAspThrAspAsnlySGlyValLeuAlaLeu 73
   |||:||||| ||| ||| |||:

```

```

131 GCTTTCAATTAG..... 144
74 GluLysAspIleAlaHisSerTyrLeuAspThrValSerLeuIleProG1 90
145 .....AGTTCAAGTTTGG 158
90 yAlaGluAsnLeuLeuSerLeuAsp...ValAspThrGluThrGlnL 106
159 TGCAGCTTGTGTAATGTTAGTTTGGTGGGTTACACAG 208
106 yslYsLeuProGluArgLysTrpAlaIleValThrSerGlySer...Pro 121
209 GTAATCTTAATATTTTGAAGACCAATAGTTGCGATCACAATTCGA 258
122 TyrLeuAlaPheSerTyrPheGluThrIleLeuLysAsnValGlyLysPr 138
259 GGCAGTCGACGACCTTAATACAGCATCAATAAATATACAAACGCG 308
138 oLysValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro. 154
309 AAATGCTTACATAGATTATATGATGTAATAATTACGTAAATAGATCCAT 358
155 .....GluGlyTyrSerArgAlaArgAspLeu 164
359 TACAATTATTTTACACACAGGTTTACA...GCAAGATATGTTTT 405
165 ArgGln.....AspLeuGlnLeuThrG1 172
406 AGACAAGTACTAAATACTATGAGGATGTTCTCAGTGCAGAGATCAGG 455
172 yLysGlnAspLeuLysTyrValIleAlaPhe...GluAspAlaProValGlyI 188
456 AAGGCTGCTTACCTTAATATATTTGGTCAACATGACATCACACATG 505
188 lElyAsAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
506 TAAAGACTGATGACAAATTTGATATAGTAGTTCGTTTAACT...AT 552
205 TyrAspLysSerValLeuPheAspAlaGlyAlaAspTyrValIleCysAs 221
553 TATGATTCACACACCTTGAGGAT..... 576
221 pLeuThrGlnValSerValIleValLysAsnAsnGluAsn 233
577 .....AAGATTGAAGAAGTTAGAACGATGCAAC 606
seq.name: /cgn2_6/prodata/2/1na/58.COMB.seq:US-08-462-679-14

seq-documentation_block:
; Sequence 14, Application US/08462679
; Patent No. 5847081
; GENERAL INFORMATION:
; APPLICANT: Michel, James L.
; APPLICANT: Kasper, Dennis L.
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Madoff, Lawrence C.
; TITLE OF INVENTION: Conjugate Vaccine For Group B
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, NW, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462.679

```

```

FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/363,311
FILING DATE: 22-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/968,866
FILING DATE: 02-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/408,036
FILING DATE: 15-SEP-1989
ATTORNEY/AGENT INFORMATION:
NAME: Bugalsky, Lawrence B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 0609,2370008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
TELEX: 248636 SSK
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1380 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: Linear
FEATURE:
NAME/KEY: CDS
LOCATION: 79..1173
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1004
OTHER INFORMATION: /note="This feature is to signify
OTHER INFORMATION: that the nucleotide sequence from position 757
OTHER INFORMATION: through 1003 is inserted at position 1004 and can be
OTHER INFORMATION: repeated up to eight times (for a total of nine
OTHER INFORMATION: repeating copies of these sequences within the
US-08-462-679-14
alignment_scores:
Quality: 109.50 Length: 229
Ratio: 0.928 Gaps: 11
Percent Similarity: 51.528 Percent Identity: 24.891
alignment_block:
US-09-558-284-2 x US-08-462-679-14 ..
Align seg 1/1 to: US-08-462-679-14 from: 1 to: 1380
27 AATATPThrLysLeuGlySerGlu...TyrGlyValAlaAspProSerGluLe 42
|||||:||||| ||| ||| ||| :|:
31 GCTATTAATTTACTGTATAGATGTAACAGATTAAGAGATATATAT 80
42 uPheLysHisSer.....HisGlyValArgThrGlnGluValLeuArgA 57
|||||:||||| ||| ||| ||| :|:
81 GTTTAAAGGTCTTAAATAATACAGTTATGATCTTACAGACAAACAC 130
57 rGPhePheProLysLeuAspAspThrAspAsnLysGlyValLeuAlaLeu 73
||||| |||
131 GCTTTCAATTAG..... 144
74 GluLysAspIleAlaHisSerTyrLeuAspThrValSerLeuIleProG1 90
145 .....AGTTCAAGTTTGG 158
90 yAlaGluAsnLeuLeuSerLeuAsp...ValAspThrGluThrGlnL 106
159 TGCAGCTTGTGTAATGTTAGTTTGGTGGGTTACACAG 208
106 yslYsLeuProGluArgLysTrpAlaIleValThrSerGlySer...Pro 121
209 GTAATCTTAATATTTTGAAGACCAATAGTTGCGATCACAATTCGA 258

```



```
122 TyrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysPro 138
    |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::
259 GGGAGTGCACGACCTTAATACAGCATCATTAATAATATACAAACGG 308
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
138 oLysValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro. 154
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
309 AAATGCTTACATAGATTATATGATGTAATAATTAGTAAATAGATCAT 358
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
155 .....GluGlyTrpSerArgAlaArgAspLeuLeu 164
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
359 TACAATTAATGTTTGTAGAACAGGTTTACA...GCAAGTATGTTTTT 405
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
165 ArgGln.....AspLeuGlnLeuThrGln 172
    |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::
406 AGACAAGCTACTAATACCTATGGGATCTTCTCAGTTGCAGCATACACG 455
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
172 yLysGlnAspLeuLysTrpValValPhe...GluAspAlaProValGlyI 188
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
456 AAGGCGTACCTTACCTATATATTTGCTGAAGATGACCTACACATG 505
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
188 IeLysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
506 TAAAGACTGCTGACAAATTGATATAGTGTGCTTAACT...ATT 552
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
205 TyrAspLysSerValLeuPheAspAlaGlyAlaAspTrpValValCysAs 221
    |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::
553 TATGATTTCACACACCTTGAGGAT..... 576
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
221 PleuThrGlnValSerValValLysAsnAsnGluAsn 233
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
577 .....AAGATTGAAGAGCTTAGAACGAATGCCAACG 606
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
```

seq_name: /cgn2_6/prodata/2/lna/5B_COMB.seq:us-08-466-210A-14

```
seq_documentation_block:
: Sequence 14, Application US/08466210A
: Patent No. 5858362
: GENERAL INFORMATION:
: APPLICANT: Michel, James L.
: APPLICANT: Kasper, Dennis L.
: APPLICANT: Ausubel, Frederick M.
: APPLICANT: Madoff, Lawrence C.
: TITLE OF INVENTION: Conjugate Vaccine For Group B
: TITLE OF INVENTION: Streptococcus
: NUMBER OF SEQUENCES: 65
: CORRESPONDENCE ADDRESS:
: ADDRESS: Sterne, Kessler, Goldstein & Fox P.L.L.C.
: STREET: 1100 New York Avenue, NW, Suite 600
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3934
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/466,210A
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/363,311
: FILING DATE: 22-DEC-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/968,866
: FILING DATE: 02-NOV-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/408,036
: FILING DATE: 15-SEP-1989
: ATTORNEY/AGENT INFORMATION:
: NAME: Bugalsky, Lawrence B.
```

```
: REGISTRATION NUMBER: 35,086
: REFERENCE/DOCKET NUMBER: 0609.237000B
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
: TELEX: 248636 SSK
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1380 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 79..1173
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 1004
: OTHER INFORMATION: /note="this feature is to signify
: OTHER INFORMATION: that the nucleotide sequence from position 757
: OTHER INFORMATION: through 1003 is inserted at position 1004 and can be
: OTHER INFORMATION: repeated up to eight times (for a total of nine
: OTHER INFORMATION: repeating copies of these sequences within the
: OTHER INFORMATION: polynucleotide)."
```

alignment_scores:

Quality:	109.50	Length:	229
Ratio:	0.928	Gaps:	11
Percent Similarity:	51.528	Percent Identity:	24.891

alignment_block:

us-09-558-284-2 x us-08-466-210A-14 ..

Align seq 1/1 to: us-08-466-210A-14 from: 1 to: 1380

```
27 AlaTrpThrLysLeuGlyTyrGlu...TyrGlyValAspProSerGluLe 42
    |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::
31 GCCTAATATTACTCTGTATAGACTTATACAGAGTAAAGGAAATATTAT 80
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
42 upheLysHisSer.....HisGlyAlaArgThrGlnGluValLeuArg 57
    |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::
81 GTTTAGAAGCTAAATAATACACTTATGATTCACAGACGAACAAAC 130
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
57 rGpPheProLysLeuAspAspThrAspAsnLysGlyValLeuAlaLeu 73
    |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::
131 GCTTTCAATTAAAG..... 144
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
74 GluLysAspIleAlaHisSerTyrLeuAspThrValSerLeuIleProGln 90
    |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::
145 .....AAGTTCAAGTTTGG 158
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
90 yAlaGluAsnLeuLeuSerLeuAsp...ValAspThrGlnThrGlnL 106
    |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::
159 TGCAGCTTCGTACTAATATGCTTAGTTTGGGGGGGTATACACAAG 208
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
106 yLysLysLeuProGluArgLysTrpAlaIleValThrSerGlySer...Pro 121
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
209 GTAATCTTAATATTGTTTGAAGAGTCAATAGTCTGCATCTACAAATCCA 258
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
122 TyrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysPro 138
    |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::
259 GCGAGTGCACGACCTTAATACAGCATCATTAATAATATACAAACGG 308
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
138 oLysValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro. 154
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
309 AAATGCTTACATAGATTATATGATGTAATAATTAGTAAATAGATCAT 358
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
155 .....GluGlyTrpSerArgAlaArgAspLeuLeu 164
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
359 TACAATTAATGTTTGTAGAACAGGTTTACA...GCAAGTATGTTTTT 405
    ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
```

```
165 ArgGln.....AspleuGlnLeuThrG1 172
|||||
406 AGACAAGGACTAAATACATGAGGATGTTTCACAGTTGACAGGTACAGG 455
172 ylysglnaspleuLysTyrValValPhe..GluaspAlaProValGlyI 188
|||||
456 AAGGCTAGCTTACTATATATATTTGGTGAAGATGACCTACACATG 505
188 lelysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSer 204
|||||
506 TAAAGCTATGACACAATTGATATAGTTAGTGTGCTTTAACT...ATT 552
205 TyrAspLysSerValLeuPheaspAlaGlyAlaAspTyrValValCys 221
|||||
553 TATGATTCAACAACCTTGAGGAT..... 576
221 pleuThrGlnValSerValValLysAsnGluAsn 233
|||||
577 .....AAGATTGAAGAATTAGAACGAATGCMAAC 606

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-467-147A-14

seq_documentation.block:
Sequence 14, Application US/08467147A
Patent No. 5908629
GENERAL INFORMATION:
APPLICANT: Michel, James L.
APPLICANT: Kasper, Dennis L.
APPLICANT: Ausubel, Frederick M.
APPLICANT: Madoff, Lawrence C.
TITLE OF INVENTION: Conjugate Vaccine For Group B
TITLE OF INVENTION: Streptococcus
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESSES:
ADDRESS: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, NW, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,147A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/363,311
FILING DATE: 22-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/968,866
FILING DATE: 02-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/408,036
FILING DATE: 15-SEP-1989
ATTORNEY/AGENT INFORMATION:
NAME: Buglisky, Lawrence B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 0609,2370009
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
TELEX: 248636 SSK
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1380 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
```

```
NAME/KEY: CDS
LOCATION: 79..1173
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1004
OTHER INFORMATION: /note="This feature is to signify
OTHER INFORMATION: that the nucleotide sequence from position 757
OTHER INFORMATION: through 1003 is inserted at position 1004 and can be
OTHER INFORMATION: repeated up to eight times (for a total of nine
OTHER INFORMATION: repeating copies of these sequences within the
US-08-467-147A-14

alignment_scores:
Quality: 109.50 Length: 229
Ratio: 0.928 Gaps: 11
Percent Similarity: 51.528 Percent Identity: 24.891

alignment_block:
US-09-558-284-2 x US-08-467-147A-14 ..

Align seg 1/1 to: US-08-467-147A-14 from: 1 to: 1380
27 ALATPThrLysLeuGlySerTyrGlu..TyrGlyValAspProSerGlu 42
|||||
31 GCCTATAATTACTCTGTATAGAGTTATACAGAGTAAAGCAATATTTAT 80
42 uPheLysHisSer.....HisGlyAlaArgThrGlnGluValLeuArg 57
|||||
81 GTTTAGAGGCTTAAATAACAGTTATGATCTTACACAGCAACCAAC 130
57 rPhePheProLysLeuAspAspThrAspAsnLysGlyValLeuAlaLeu 73
|||||
131 GCTTTCAATTAAAG..... 144
74 GluLysAspIleAlaHisSerTyrLeuAspThrValSerLeuIleProG1 90
|||||
145 .....AAGTTCAAGTTGG 158
90 yAlaGlnAsnLeuLeuSerLeuAsp..ValAspThrGlnThrGln 106
|||||
159 TGCAGCTCTGTACTAATTGCTGTAGTTTGGTGGGTTACACAAAG 208
106 yLysValPheProGluArgLysTrpAlaIleValThrSerGlySer...Pro 121
|||||
209 GTATCTTATATTTTGAAGAGTCAATAGTTCGATCTACAAATCCA 258
122 TyrLeuAlaPheSerTyrPheGlnThrIleLeuLysAsnValGlyLysPr 138
|||||
259 GGGAGTGCAGCGACCTTAATACAAACATCATAAATAATACAAACCG 308
138 ouLysValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
|||||
309 AATGCTTACATAGATTATATGATGTAATAATTGTAATAAAGATCCAT 358
155 .....GluGlyTyrSerArgAlaArgAspLeuLeu 164
|||||
359 TACAATTAATTGTTTAGAACAAAGTTTACA...GCAAGATATGTTT 405
165 ArgGln.....AspleuGlnLeuThrG1 172
|||||
406 AGACAAGGACTAAATACATGAGGATGTTTTCACAGTTGACAGGTACAGG 455
172 ylysglnaspleuLysTyrValValPhe..GluaspAlaProValGlyI 188
|||||
456 AAGGCTAGCTTACTATATATATTTGGTGAAGATGACCTACACATG 505
188 lelysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSer 204
|||||
506 TAAAGCTATGACACAATTGATATAGTTAGTGTGCTTTAACT...ATT 552
205 TyrAspLysSerValLeuPheaspAlaGlyAlaAspTyrValValCys 221
```

|||||.....|||
553 TATGATTCACACACCTTGACGAT..... 576
221 pleuThrcInvalSerValValysAsnAsnGluAsn 233
577AAGATTGAAGAGATTAGACGAATGCCAAC 606
seq_name: /cgn2_6/prodata/2/lna/5B_COMB.seq:US-08-469-014-14

seq_documentation_block:
; Sequence 14, Application US/08469014
; Patent No. 5968521
; GENERAL INFORMATION:
; APPLICANT: Michel, James L.
; APPLICANT: Kasper, Dennis L.
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Madoff, Lawrence C.
; TITLE OF INVENTION: Conjugate Vaccine Against Group B
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, NW, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,014
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/363,311
; FILING DATE: 22-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/968,866
; FILING DATE: 02-NOV-1992
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/408,036
; FILING DATE: 15-SEP-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Bugalsky, Lawrence B.
; REGISTRATION NUMBER: 35,086
; REFERENCE/DOCKET NUMBER: 0609,2370006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; TELEX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1380 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 79..1173
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1004
; OTHER INFORMATION: /note="this feature is to signify
; OTHER INFORMATION: that the nucleotide sequence from position 757
; OTHER INFORMATION: through 1003 is inserted at position 1004 and can be
; OTHER INFORMATION: repeated up to eight times (for a total of nine
; OTHER INFORMATION: repeating copies of these sequences within the
; OTHER INFORMATION: polynucleotide)."
US-08-469-014-14

alignment_scores:
Quality: 109.50 Length: 229
Ratio: 0.928 Gaps: 11
Percent Similarity: 51.528 Percent Identity: 24.891

alignment_block:
US-09-558-284-2 x US-08-469-014-14 ..

Align seq 1/1 to: US-08-469-014-14 from: 1 to: 1380

27 AlatrThrLysLeuCysTyrGlu...TyrGlyValAspProSerGluLe 42
|||||.....||||| 111 111
31 GCCTAATATTACCTCTGATAGAGTTATACAGAGTAAGAGCAATATTAT 80
42 uPheLysHisSer.....HisGlyAlaArgThrGlnGluValLeuAla 57
|||||.....| 111:.....|
81 GTTTAGAAAGGTCTAAATAATACAGTTATGATCTTCACAGACGAACAC 130
57 rPhePheProLysLeuAspAspThrAspAsnLysGlyValLeuAlaLeu 73
||||| 111
131 GTTTTCATTAG..... 144
74 GluLysAspLleAlaHisSerTyrLeuAspThrValSerLeuLleProG 90
145AAGTTCAAGTTTGG 158
90 ValAlaGluAsnLeuLeuLeuSerLeuAsp...ValAspThrGluThrGln 106
|||||| 111
159 TGCACCTTCTGTACTAATTTGGCTTAGTTTGGGCGGTACACAG 208
106 yLysLeuProGluArgLysTrpAlaIleValThrSerGlySer...Pro 121
||| 111
209 GTAACTTAATATTGTTTGAAGCATAGTTCGTCATCTACAAATTGCA 258
122 TyrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysPr 138
||| 111
259 GCGAGTCGACGACCTTAATATACAGCATCACTAAATAATATCAAAACGG 308
138 OlyValAlaPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
||| 111
309 AAATGCTTACATGATTTATGATGATGATAAATTAGTAAATATAGTCAT 358
155GluGlyTyrSerArgAlaArgAspLeuLeu 164
||| 111
359 TACAACTTAATTGTTTACAGACAGGTTTAC...GCCAAGATGTTT 405
165 ArgGln.....AspLeuGlnLeuThrG 172
|||||| 111
406 AGACAAGCTACTAATTAATGCGGATGTTTCTCAGTTGACAGATACAGG 455
172 yLysGlnAspLeuLysTyrValValPhe...GluAspAlaProValGlyI 188
||| 111
456 AAGGCTAGTCTTACCTTAATATATTTGGTGAAGATGAGTACACATG 505
188 IeLysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
||| 111
506 TAAAGACTGATGACAAATTTGATATAGTTAGTTGCTTAACT...ATT 552
205 TyrAspLysSerValLeuPheAspAlaGlyAlaAspTyrValValCysAs 221
|||||.....| 111
553 TATGATTCACACACCTTGACGAT..... 576
221 pleuThrcInvalSerValValysAsnAsnGluAsn 233
577AAGATTGAAGAGATTAGACGAATGCCAAC 606
seq_name: /cgn2_6/prodata/2/lna/PCTUS_COMB.seq:PCT-US93-10506A-14
seq_documentation_block:
; Sequence 14, Application PC/TUS9310506A
; GENERAL INFORMATION:
; APPLICANT: THE GENERAL HOSPITAL CORPORATION

```

APPLICANT: Fruit Street
APPLICANT: Boston, Massachusetts 02114
APPLICANT: United States of America
APPLICANT: 75 Francis Street
APPLICANT: Boston, Massachusetts 02115
APPLICANT: United States of America
TITLE OF INVENTION: Conjugate Vaccine Against Group B
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/10506A
FILING DATE: 02-NOV-1993
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/966,866
FILING DATE: 02-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Cimbalia, Michele A.
REGISTRATION NUMBER: 33,851
REFERENCE/DOCKET NUMBER: 0609,237PC01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
TELEX: 248636 SSK
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1380 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 79..1173
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1004
OTHER INFORMATION: /note="This feature is to signify
OTHER INFORMATION: that the nucleotide sequence from position 757
OTHER INFORMATION: through 1003 is inserted at position 1004 and can be
OTHER INFORMATION: repeated up to eight times (for a total of nine
OTHER INFORMATION: repeating copies of these sequences within the polynucleotide
PCT-US93-10506A-14

alignment_scores:
Quality: 109.50 Length: 229
Ratio: 0.928 Gaps: 11
Percent Similarity: 51.528 Percent Identity: 24.891

alignment_block:
US-09-558-284-2 x PCT-US93-10506A-14 ..
Align seg 1/1 to: PCT-US93-10506A-14 from: 1 to: 1380

27 AlATpThLysLeuCySTYrgJu...TyrgLyValAspProSerGlue 42
|||||: ||||| ||| ||| ||| : : :
31 GCCTAAATTTACTGTGTATAGAGTTATACAGAGTAAGAGAAATTTAT 80
|||||: ||||| ||| ||| ||| : : :
42 ubpelyshLsSer.....HisGlyAlaArGthrGInGluValLeuArg 57
|||||: ||||| ||| ||| ||| : : :
81 GTTTAGAGGTCTAAAAATAACAGTTATGATCTTACACAGCAACAAAC 130

```

```

57 rGpHePheProLysLeuAspThrAspAsnLysGlyValLeuAlaLeu 73
|||||: ||||| ||| ||| ||| : : :
131 GGTITTTCAATTAAAG..... 144
74 GluLysAspLLeaLanISerTYrLeuAspThrValSerLeuIleProG 90
|||||: ||||| ||| ||| ||| : : :
145 .....AAGTTCAGGTTTGG 158
90 yAlAGLuanleuLeuSerLeuAsp...ValAspThrGInGlu 106
|||||: ||||| ||| ||| ||| : : :
159 TGCACCTTCTGACTAATTGGCTTACTTTTGGTGCGGTACACAG 208
|||||: ||||| ||| ||| ||| : : :
209 GTAACTTAATATTTTGAAGCATAGTGTGCATCTACATTCACA 258
|||||: ||||| ||| ||| ||| : : :
106 ySLysLeuProGluArGlySTrPalLeValThrSerGlySer...Pro 121
|||||: ||||| ||| ||| ||| : : :
122 TyrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysPr 138
|||||: ||||| ||| ||| ||| : : :
259 GGGAGTCGACGCGCCTTAATAACAGCATCTAAATAATACAAACGG 308
|||||: ||||| ||| ||| ||| : : :
138 OLSValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
|||||: ||||| ||| ||| ||| : : :
309 AATGCTTACATGATTATATGATGCTAAATTAGCTAAATAGATCCAT 358
|||||: ||||| ||| ||| ||| : : :
155 .....GluGlyTYrSerArgAlaArgAspLeu 164
|||||: ||||| ||| ||| ||| : : :
359 TACAAATTATTTTGAACAGGTTTACA...GCAAGCATGTTT 405
|||||: ||||| ||| ||| ||| : : :
165 ArgGln.....AspLeuGlnLeuThrG 172
|||||: ||||| ||| ||| ||| : : :
406 AGACAAGGTACTAAATACTAGGAGTGTTCAGTTCCAGTACAGACG 455
|||||: ||||| ||| ||| ||| : : :
172 yLysGlnAspLeuLysTYrValValPhe...GluAspAlaProValGly 188
|||||: ||||| ||| ||| ||| : : :
456 AAGGCTACTCTTACTATATATATTTGGTGAAGTGACATCCACANG 505
|||||: ||||| ||| ||| ||| : : :
188 LeuSAlaGlyLysAlaMetGlyAlaIleThrValGlyLysThrSerSer 204
|||||: ||||| ||| ||| ||| : : :
506 TAAAGACTGATGACAAATGATATAGTAGTGTCTTAACT...ATT 552
|||||: ||||| ||| ||| ||| : : :
205 TyrAspLysSerValLeuPheAspAlaGlyAlaAspTYrValValCysAs 221
|||||: ||||| ||| ||| ||| : : :
553 TATGATTCAACAACCTTGAGGAT..... 576
221 pleuThrGlnValSerValValLysAsnGluAsn 233
|||||: ||||| ||| ||| ||| : : :
577 .....AAGATTGAGAGATTAGAACGAATGCAAC 606

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-10506-14
seq_documentation_block:
Sequence 14, Application PC/TUS9310506
GENERAL INFORMATION:
APPLICANT: Michel, James L.
APPLICANT: Kasper, Dennis
APPLICANT: Ausubel, Frederick M.
APPLICANT: Madoff, Lawrence C.
TITLE OF INVENTION: Conjugate Vaccine Against Group B
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/10506
FILING DATE: 02-NOV-1993
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/968,866
FILING DATE: 02-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Cimbalia, Michele A.
REGISTRATION NUMBER: 33,851
REFERENCE/DOCKET NUMBER: 0609.237PC01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
TELEX: 248636 SSK
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1380 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 79..1173
NAME/KEY: misc_feature
LOCATION: 1004
OTHER INFORMATION: /note="this feature is to signify
OTHER INFORMATION: that the nucleotide sequence from position 757
OTHER INFORMATION: through 1003 is inserted at position 1004 and can be
OTHER INFORMATION: repeated up to eight times (for a total of nine
OTHER INFORMATION: repeating copies of these sequences within the polynucleotide
PCT-US93-10506-14

alignment_scores:
Quality: 109.50 Length: 229
Ratio: 0.928 Gaps: 11
Percent Similarity: 51.528 Percent Identity: 24.891

alignment_block:
US-09-558-284-2 x PCT-US93-10506-14 ..
Align seg 1/1 to: PCT-US93-10506-14 from: 1 to: 1380

27 Alatrphtrhrylsleucystyrglu...TyrGlyAlaAspProSerGluLe 42
|||||:|||||:|||||:|||||:|||||:
31 GCCATATATTACTCTGTATAGAGTTATACAGTAAGCAAGATATTAT 80
42 uphelyshisler.....HisGlyAlaArgThrGlnGluValLeuArg 57
|||||:|||||:|||||:|||||:|||||:
81 GTTAGAGAGGCTAAATAAATAACAGTTATGACTTCACAGCAAGAACAC 130
57 rghperhpepPolysleuAspAspThrAspAsnLysGlyValLeuAlaLeu 73
|||||:|||||:|||||:|||||:|||||:
131 GGTTCATTAATAAG.....AAAGTTCAAGCTTGG 144
74 GlulysAspIleAlaHisSerTyrLeuAspThrValSerIleProGlu 90
|||||:|||||:|||||:|||||:|||||:
145 .....AAAGTTCAAGCTTGG 158
90 yAlaGluAsnLeuLeuSerLeuAsp...ValAspThrGluThrGlnL 106
|||||:|||||:|||||:|||||:|||||:
159 TGCAGCTCTGTACTAATGTGCTTACTTTTGGTGGGATTCACACAG 208
106 yslYsleuProGluArgLysTrpAlaIleValThrSerGlySer...Pro 121
|||||:|||||:|||||:|||||:|||||:
209 GTAATCTTAATATTATTTTGAAGAGCAATAGTCTGCATCTACAAATTC 258
122 TyrLeuAlaHisSerTrpPheGluThrIleLeuLysAsnValGlyLys 138
|||||:|||||:|||||:|||||:|||||:
259 GGGAGTGCAGCGACCTTAATATACAGCATCAATAAATATACAAACCG 308

```

```

138 OlyValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro. 154
|||||:|||||:|||||:|||||:|||||:
309 AATGCTTACATAGATTATATGATGTAAATAGTAATATAGATTCAT 358
155 .....GluGlyTyrSerArgAlaArgAspLeuLeu 164
|||||:|||||:|||||:|||||:|||||:
359 TACATTAAATTGTTTAGACACAGCTTTTACA...GCCAAGTATGTTTT 405
165 ArgGln.....AspLeuGlnLeuThrGlu 172
|||||:|||||:|||||:|||||:|||||:
406 AGACAGCTACTAAATTAATGAGGATGTTTCTCATGTCAGAGTACAGG 455
172 yLysGlnAspLeuLysTyrValValPhe...GluAspAlaProValGly 188
|||||:|||||:|||||:|||||:|||||:
456 AAGGCTAGTCTTACCTATATATATTTGGTGAAGATGAGCTACCAATG 505
188 leLysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
|||||:|||||:|||||:|||||:|||||:
506 TAAAGACTGATGACCAAAATGTATATAGTTAGTTGCTTTAACT...ATT 552
205 TyrAspLysSerValLeuPheAspAlaGlyAlaAspTyrValValCysAs 221
|||||:|||||:|||||:|||||:|||||:
553 TATGATTCAACACACTTGAGGAT..... 576
221 pleuThrGlnValSerValValLysAsnAsnGluAsn 233
|||||:|||||:|||||:|||||:|||||:
577 .....AAGATTGACAGAAAGTTAGACGATGCAAC 606

seq.name: /cgn2_6/prodata/2/ina/5h_comb.seq:US-08-937-102-1
seq.documentat_block:
Sequence 1, Application US/08937102A
Patent No. 5965134
GENERAL INFORMATION:
APPLICANT: Thiel, Heinz-Jurgen
APPLICANT: Elpers, Knut
APPLICANT: Pauly, Thomas
TITLE OF INVENTION: T Cell Stimulating Protein of Pestivirus
FILE REFERENCE: 1/94108
CURRENT APPLICATION NUMBER: US/08/937,102A
CURRENT FILING DATE: 1997-09-24
EARLIER APPLICATION NUMBER: 08/693,247
EARLIER FILING DATE: 1996-08-16
NUMBER OF SEQ ID NOS: 36
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 3639
TYPE: DNA
ORGANISM: Classical Swine Fever Virus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(3639)
US-08-937-102-1

alignment_scores:
Quality: 91.00 Length: 300
Ratio: 0.599 Gaps: 21
Percent Similarity: 50.667 Percent Identity: 24.333

alignment_block:
US-09-558-284-2 x US-08-937-102-1 ..
Align seg 1/1 to: US-08-937-102-1 from: 1 to: 3639

7 AspleuCySleuPheAspLeuAspGlyThrIleValSerThrThrValAl 23
|||||:|||||:|||||:|||||:|||||:
2253 AACCTATGCCCTCTTACGGTTA.....CTTTGCCAGATGTACAAACCA 2296
23 alaGluLys.....AlatrphtrL 30
|||||:|||||:|||||:|||||:|||||:
2297 AGCTGAGCGCGCAATGCTAGATATTCCTTTATATTCTAGATGAGTAT 2346

```

```
30 ysLeuCysTyr.....GluTyrGlyValAspProSerGlu 41
   :|||||
2347 CATTGTGCTACCCAGAACAACTGCGCAATCATGTGGGAAGATCCACAGAT 2396
   ||| |||
42 LeuPheLysHisSerHisGlyAlaArgThrGluValLeuArgArgPh 58
   ||| |||
2397 CTCGAGAAACCTGCGGTGTAGCTATGACCGCACCGCGAGCGACAG 2446
   :|||
58 ePheProLysLeuAspThrAspAsnLysGlyValLeuAlaLeuGlu 74
   :|||
2447 TAACAACCACTGGGCAAAACACCTATAGAGAAATTATAGCCCCGAA 2496
   :|||
75 .....LysAspIleAlaHisSerTyrLeuAspThrValSerLe 87
   :|||
2497 GTGATGAAGAGAGAAAGACTTGGGTCTGAGTACTAGATATTGCCGACT 2546
   :|||
87 u...IleProGlyValGlu.....AsnLeuLeu.....S 97
   :|||
2547 GAAGATACCGATAGAGAGATGAAGAATACATCTAGTTTGTGCCCA 2596
   :|||
97 erLeuAspValAspThrGluThrGlnLysLysLeuProGluArgLysTrp 113
   :|||
2597 CCAAGAACATGCGCGGTAGAGCGGCAAGAAATTGAAGCCCAAGCATAC 2646
   :|||
114 .....AlaIleVal 116
   :|||
2647 AACTCGGGCTACTACTACAGCGAGAGACCCATCTAAGTACGAGGTGT 2696
   :|||
116 lThrSerGlySerProTyrLeuAla..... 124
   :|||
2697 GACGTCCGACGTCCCATACGTGGTGTAGCAACCAAGCATAGATCGG 2746
   :|||
125 .....PheSerTrpPheGluThrIleLeuLysAsnVal 135
   :|||
2747 GCCTTACCCCTCCGACCTGACGCGTGTGTGACACGCGACTCAAGTGT 2796
   :|||
135 lGlyLys.....ProLysValPheIleThrGlyPheAsp 146
   :|||
2797 GAAAAAAGAAATCCGACTGTCAACCCAGATGCTTCATATGACGCGCCT 2846
   :|||
147 ValLysAsnGlyLys.....ProAspPro...GluG1 156
   :|||
2847 GAAAGAAATGCGCCGTCACTATTGGGAAACAGCCAGAAAGAGCGAGG 2896
   :|||
156 yTyr...SerArgAlaArgAspLeuLeuArgGlnAspLeuGlnLeuThr 171
   :|||
2897 TTGGAAAGAGTGAAGCCGCGGAGATACTACAGGAGTCAAGAAACACCTGTC 2946
   :|||
172 GlyLysGlnAspLeuLysTyrValValPheGluAspAlaProValGly11 188
   :|||
2947 GGCTCTAAGAACTACCATTTATGCTTATGCAAGCCAGAGGTACGCGCAT 2996
   :|||
188 eLysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
   :|||
2997 AGAAGATGCGG.....ATAAATATCACCAATCCT 3025
   :|||
205 .....TyrAspLysSerValLeuPheAspAlaGlyAlaAsp 216
   :|||
3026 TCAGAGAGATGACTACGACTGAGGCTTTATGAGGAA.....GAT 3066
   :|||
217 TyrValValCysAspLeuThrGlnValSerValValLysAsn 230
   :|||
3067 AGCCTGATG.....ATCAGACACTGAAATCCTCAACAC 3102
```



```

VERSION      A2683512.1  GI:11820658
KEYWORDS     GSS.
SOURCE       Entamoeba histolytica.
ORGANISM     Entamoeba histolytica.
              Eukaryota; Entamoebidae; Entamoeba.
REFERENCE    1 (bases 1 to 921)
              Loftus,B., Van Aken,S. and Fraser,C.
AUTHORS      Determination of clone end sequences from Entamoeba histolytica
              HMI:IMSS sheared DNA library
              Unpublished (2000)
JOURNAL      Contact: Brendan J Loftus
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: b.loftus@tigr.org
COMMENT      Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
              DNA library
              Seq primer: M13-Forward
              Class: shotgun
              High quality sequence start: 29
              High quality sequence stop: 769.
              Location/Qualifiers
FEATURES
  source
    1..921
    /organism="Entamoeba histolytica"
    /strain="HMI:IMSS"
    /db_xref="taxon:5759"
    /clone_lib="Entamoeba histolytica Sheared DNA"
    /note="Vector: pHOSt1; Site_1: Bst I; Constructed at the
    Institute for Genomic Research (TIGR), Rockville, MD.
    Genomic DNA isolated from broth cultures of E. histolytica
    using a method described by Clark and Diamond (Clark,
    C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
    method for isolate identification. Exp. Parasitol.
    77:450.). The DNA was mechanically sheared to give a
    tight size distribution (~2 kb). The v + i method used for
    the library construction is described in detail in Smith,
    H.O. and Venter, J.C. (Making small insert libraries for
    whole genome shotgun sequencing projects. in Genome
    Sequencing: A Practical Approach, eds. M. Vaudin and B.
    Barrell, Oxford University Press, 1999)."
```

```

BASE COUNT   291 a      134 c      125 g      371 t
ORIGIN
alignment_scores:
  Quality: 150.00      Length: 225
  Ratio: 1.339         Gaps: 10
  Percent Similarity: 49.778      Percent Identity: 26.222
```

```

alignment_block:
US-09-558-284-2 x A2683512/rev ..
```

```

Align seg 1/1 to reverse of: A2683512 from: 1 to: 921
```

```

10 leupheaspheusapgythrrilevalserththrvalalalaguly 26
   ::::::::::::::::::::::::::::::::::::::::::::|||
631 ATCTTTCATCTTGATGGACACTACTTGATGACA.....GAAAT 594
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|
26 salatrpthrlyleucystyrglylvalaspprosergluueup 43
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|
593 ATTAATATACAAATGTTTACACACATAT.....TTAG 562
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|
43 helyshisSerHsGlyAlaAarg...ThrlngluvalleuAargAphe 58
   ::::::::::::::|::|::|::|::|::|::|::|::|::|
561 ATGAATATGCTAATGAGCTAAATTTACTATTAATAAAGAAATG 512
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|
59 pheprolysleuaspapthraspasnlygllyalleu.....71
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|
511 ATGGGAAGACATATGAGTTGCTACTTAATGTTAATGATATTTTCA 462
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|
72 .....Alaleugluysaspilaleahis....79
```

```

461 TATTAATGACACACTGGAACATGCATACATATAAATTCATTTAA 412
   |||::|::|::|::|::|::|::|::|::|::|::|::|
80 ..SerTyrleuaspthrvalserleuileproglialagluasnleu 95
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
411 ATTAATATGCTAATGAGCTAATAACCATTCACAGACATGAAGATTVA 362
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
96 .....leuserleuaspvalaspthrcluhtrgl 105
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
361 AATTATTTTAAAAAACATCACAATTCCTATAGCTCTGCTACTTCAACAC 312
   |||::|::|::|::|::|::|::|::|::|::|::|::|
105 nlvslvleuaprogluarglystrp.....113
   |||::|::|::|::|::|::|::|::|::|::|::|::|
311 TAAATGCTGCTTTGAACCAAAAAATGTTAAAAACCAAGATGCTCAATT 262
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
114 .....AlallevalthrSerglySerprotyrleuAlapheserTrp 127
   |||::|::|::|::|::|::|::|::|::|::|::|::|
261 ATTTGATGCAATTTGTTCTTGAGATGATCCCTCAT.....227
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
128 phegluThrileuEulysnvalglylvsprolyvalphelethrgl 144
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
227 .....227
144 ypheaspvalysasnlylvsproaspprogluyltyrSeraArgAla 161
   |||::|::|::|::|::|::|::|::|::|::|::|::|
226 .....GTAAAGAGCTAAACCAATCCACAATATTTTTCATGCTG 184
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
161 rgaAspleuEuarvgluAspleucluhleuthrlylvsSglnaspleuys 177
   |||::|::|::|::|::|::|::|::|::|::|::|::|
183 GTCATTACTT.....GGATGACTGACATGAAA 155
   |||::|::|::|::|::|::|::|::|::|::|::|::|
178 Tyr...ValValphegluAspAlaprovalglylleysAlaglysal 193
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
154 CAAGCAATTTATTTGAAGTCAGTACTTGCTGTTCAACAGGAATTGC 105
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
193 ametGlyAlallethrvalglylle 201
   |::|::|::|::|::|::|::|::|::|::|::|::|
104 ATCTGAGCTTATATCTAGTACTTAT 80
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
seq_name: gb_est17:AI209756
seq_documentation_block:
LOCUS      AI209756      266 bp      mRNA      19-OCT-1998
DEFINITION c9b02a1.r1 Aspergillus nidulans 24hr asexual developmental and
              vegetative cDNA lambda zap library Aspergillus nidulans cDNA clone
              c9b02a1 5', mRNA sequence.
ACCESSION  AI209756
VERSION    AI209756.1  GI:3771698
KEYWORDS   EST.
SOURCE     Aspergillus nidulans.
ORGANISM   Aspergillus nidulans.
              Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
              Eurotiales; Trichocomaceae; Emericella.
REFERENCE  1 (bases 1 to 266)
              Kupfer,D., Gray,D., Hausner,J., Lai,H., Martin,W., Aramayo,R.,
              Prade,R. and Roe,B.
              An Aspergillus nidulans EST Database
              Unpublished (1998)
              Other_ESTs: c9b02a1.f1
              Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
              Department of Chemistry and Biochemistry
              Advanced Center for Genome Technology, University of Oklahoma
              620 Parrington Oval, Norman, OK 73019, USA
              Tel: 405 325 4912
              Fax: 405 325 7762
              Email: broe@ou.edu
              We anticipate the future release of the cDNA clones to the Fungal
              Genetics Stock Center
              Seq primer: 73
              High quality sequence stop: 161.
              Location/Qualifiers
FEATURES
  source
    1..266
    /organism="Aspergillus nidulans"
```



```

/strain="EGSC_A26"
/db_xref="taxon:5072"
/clone="c9b02a1"
/clone_lib="Aspergillus nidulans 24hr asexual
developmental and vegetative cDNA lambda zap library"
/lssue_type="vegetative mycelia, asexual structures"
/Note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
3' end of cDNA cloned into XhoI site of pBluescript"
BASE COUNT      56 a      66 c      84 g      60 t
ORIGIN
alignment_scores:
      Quality: 139.00      Length: 97
      Ratio: 2.242      Gaps: 3
Percent Similarity: 63.918      Percent Identity: 35.052
alignment_block:
US-09-558-284-2 x AI209756 ..
Align seg 1/1 to: AI209756 from: 1 to: 266
113 TTPAAlAlleValThSerGlySerProTyrLeuAlaAlaPheSerTyrPheG1 129
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 TGGGGGGCTTGTGACGTCGGGAGCCGGCCGTCATTGACCGCGTGGCGG 50
129 uThrLeuLysAsnValG1yLysProLysValPheLeuThrG1yPheA 146
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 TGTTC TC.....AACTTCACACATTCACATGTCCTGCTGTCGGAGG 94
146 sPValLysAsnG1yLysProAspProG1uG1yTyrSerArgAlaArgAsp 162
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
95 ATCTACAACTCGCGGAGCCAGATCGCGATGTCACCTGCTGGACGAAG 144
163 LeuLeuArgG1nAspLeuG1nLeuThrG1yLysG1nAspLeuLysTyr.. 178
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
145 .....AAATGGGCTGTGAAACATTTC 164
179 .....ValValPheG1uAspAlaProValG1yIleLysAlaG1yL 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
165 CTCCTCGATAGTGGTCTCTGGAGGACGCCCTTCAGGCAATAAAGCTGGCA 214
192 ySalMetG1yAlaIleThrValG1yIleThrSerSerTyr 205
||||| ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
215 AGCGACCGCTTTCACACTGTCATTGCTTTCACCACTGAC 255
seq_name: gb_estr7:BG129500
seq_documentation_block:
LOCUS      BG129500      717 bp      mRNA      EST      31-JAN-2001
DEFINITION      EST1475146 tomato shoot/meristem Lycopersicon esculentum cDNA clone
cDNA2P11 5' sequence, mRNA sequence.
ACCESSION      BG129500
VERSION      BG129500.1 GI:12629688
KEYWORDS      EST.
SOURCE      tomato.
ORGANISM      Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 717)
van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Uterback,T.,
Hansen,C., Renning,C. and Tanksley,S.
Generation of ESTs from tomato shoot/meristem tissue
Unpublished (2001)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.
Location/Qualifiers
FEATURES

```

```

source
1. 717
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CFOF24P11"
/clone_lib="tomato shoot/meristem"
/tissue_type="shoot/meristem"
/dev_stage="developing shoots from 4-6wks old plants"
/lab_host="SOLR"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2: XhoI; Small expanding leaves from the growing tip were taken from greenhouse plants (4-6wks old TA496). Tissue was immediately frozen in liquid nitrogen."

BASE COUNT      208 a      129 c      177 g      203 t

ORIGIN

alignment_scores:
      Quality: 139.00      Length: 205
      Ratio: 1.311      Gaps: 7
Percent similarity: 51.707      Percent identity: 27.317

alignment block:
US-09-558-284-2 x BG129500 ..

Align seg 1/1 to: BG129500 from: 1 to: 717

10 LeuPheAspLeuAspGlyThrIleValSerThrThrValAlaAlaGly 26
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
69 CTrATTGATGATGACCGAAGACTTTATGTGATTAGATCCCTTCACTACTA 118
26 salATPrThrLysLeuCySTyGrluTyGly.....V 37
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
119 TGCCTTCGCGTAATGCTTCTAGACATAGCAATGACGGCTGGCCAG 168
37 aLAspProSerGluLeuPheLysHisSerHisGlyAlaArgThrGlnGlu 53
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
169 TTGATGAGGAGTGCGTTTATCAAGACTATTGCGGAACACACACATGAT 218
54 ValLeuArgArgPhePheProLysLeuAspSerThrAspAsnLysGly 70
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
219 ATTGCTTGGCCCTTTTCTCT.....GATGATCGAAGACGGGCTTTAA 262
70 IleuAlaLeuGluLysAspIleAlaHisSerTrpLeuAspThrValSerL 87
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
263 ATTTTGTGATGAAGAAGACGTATGTTGAAGCTTGGTGAAGAACACT 312
87 euLleProGlyAlaGluAsnLeuLeuSerLeuAspValAspThrGlu 103
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
313 TCGAACCT.....ATGATGGTTTGTACAG 338
104 ThrGlnLysLysLeuProGluArg.....LysTrpAlaIleValThrSe 118
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
339 GTGAAGAAATTTGATTGAAGATCGTGACTGAACGGCTGCATTAACA 388
118 rGlySerProTyLeuAlaPheSerTrpPheGluThrIleLeuLysAsnV 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
389 TGCACCTACACTAAACGCT.....CAACTGATTAATAGAAATAC 426
135 aLGlYLySProLysValPhe.....IleThrGlyPheAspValLys 148
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
427 TTGGCTTAAGAGATTCTTTGATGTAGTATTATATCGAAGACAGTGAG 476
149 AsnGlyLysProAspProGluGlyTrpSerArgAlaIleAspLeuLeu 165
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
477 CGGGCAAAACATCTCCGTACCTTACTTGAAGGCGCTTGAACCTCTTAA 526
165 gLnAspLeuGlnLeuThrGlyLysGlnAspLeuLysTyTrpValPheG 182
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
527 CGGTCTTAAGAGACACAG.....TTCAATATTGG 555
182 LuAspAlaProValGlyIleLysAlaGlyLysAlaMetGlyAlaIleThr 198
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

556 AGGATTCGTCCTCCGGAAATAAACGTGGGGTTCGACGCGGATCCTGTT 605
199 ValGlyIleThrSer 203
      |||||.....
606 GTTGGTTGGCAACT 620
seq_name: gb_gss28:AZ543635

```

LOCUS	A2543635	903 bp	DNA	GSS	14-NOV-2000
DEFINITION	ENTHE27TF Entamoeba histolytica Sheared DNA. Entamoeba histolytica genomic, DNA sequence.				
ACCESSION	A2543635				
VERSION	A2543635				
KEYWORDS	GSS.				
SOURCE	Entamoeba histolytica.				
ORGANISM	Entamoeba histolytica.				

REFERENCE	1 (bases 1 to 903)
AUTHORS	Loftus,B., Van Aken,S. and Fraser,C.
TITLE	Determination of clone end sequences from <i>Entamoeba histolytica</i>
JOURNAL	HMI:IMSS sheared DNA library
COMMENT	Unpublished (2000)
	Contact: Brendan J Loftus
	Department of Microbiology, University of

BASE COUNT	329 a	95 c	146 g	333
ORIGIN				

```
alignment_scores:
  Quality: 131.00      Length: 248
  Ratio: 1.000         Gaps: 12
Percent Similarity: 52.823  Percent Identity: 23.790
```

alignment_block:
US-09-558-284-2 x AZ543635

Align seg 1/1 to: AZ543635 from: 1 to: 903

10 leupheapspleuaspglythrilevalserthrthrvalalaalagluly 26
 |||||:|||||:|
 149 ttgtttgatttttaattggaccttttaattgtaaacctcttcattgcatt 198

```

26  salatprhrlyslencystyrglytyclvalaspro.....Serg 41
    ||||| ..... |||
199  TTGTGTGAAAGAAATGGCAAAAAGATAAGAGTACTCCATTAAAGTGAG 248
    ||||| ..... |||
41  IuLeuPheLysHisSerHisGlyAlaArgThrGlnIuValLeuArg 57
    ||||| ..... |||
249  ATGATTTAATAATTATTGAATGAGTGAACAATAAATCAATTAATTGAACAT 298
    ||||| ..... |||
58  phePheProLys..LeuAspAspThrAspAsnLysGlyValLeuAla 73
    ||||| ..... |||
299  ATTCTCAATTAAGAAATAGTAGATGACATGCACAAAGAAATATGCTCAACA 348
    ||||| ..... |||
73  uGluLysAspLLeaLanHisSerTyrLeu.....AspThrValSerLeu 88
    ||||| ..... |||
349  AAAAGCAAAATTATATATAGCAACTATGTTAATGAAGAACTGATATAAATTAT 398
    ||||| ..... |||
88  LeProGlyValaGluAsnLeuLeuLeuSerLeuAspValAspThrGluThr 104
    ||||| ..... |||
399  GTGATGGTGCCCAATTATATCTTTT.....GAGGCA 427
    ||||| ..... |||
105  GlnLysLys.....LeuProGluArgLysTrpAlaIleValThrSerg 119
    ||||| ..... |||
428  TTAAAAAAATGTATATATCTCT.....TTTCAATTGGCCACTTCAAG 468
    ||||| ..... |||
119  Y.....SerProTyrLeuAlaPheSert 127
    ||||| ..... |||
469  TCATTGGGGAATGTTCAAGCTGTTTATTCAAAAATACCACTTGGAGAAAT 518
    ||||| ..... |||
127  rpPheLuhThrIleLeuLysAsnValGlyLysProLysValPheIleThr 143
    ||||| ..... |||
519  GGTTTGATATTT.....GATAAAATATATTTTAAAT 547
    ||||| ..... |||
144  GlyPheAspValLysAsnGlyLysProAspProGluGlyTyrSerArg 160
    ||||| ..... |||
548  GATTTACCTTTTAA...GCTAAACCTGCCTCGACATTTATTTAAAGC 594
    ||||| ..... |||
160  aArgAspLeuLeuArgGlnAspLeuGlnLeuThrGlnLysGlnAspLeu 177
    ||||| ..... |||
595  TTCTTAAGAAATTA.....GGAGTTAGTATTTCTC 623
    ||||| ..... |||
177  ystYrValValPheGluAspAlaProValGlyIleLysAlaGlyLysAla 193
    ||||| ..... |||
624  ATGTGATTTGATTTGAAGACTATTTACAGTATTCATCTCGCATTTATCA 673
    ||||| ..... |||
194  MetGlyAlaIleThrValGlyIleThrSerSertYrAspLysSerValLe 210
    ||||| ..... |||
674  GCTGGTGCTACTCTCCAAATAGCAATTCCTCTCGAAATGACTGTTAATGAT 723
    ||||| ..... |||
210  uPheAsp...AlaGlyAlaAspTyrValValLysAspLeuThrGlnValS 226
    ||||| ..... |||
724  ACTTCAAAATTAAGAGATGCACCTGCAATTCACACATTTAATAGATAAA 773
    ||||| ..... |||
226  eYrVal.....ValLysAsnAsnGluAsnGly 234
    ||||| ..... |||
774  CAATTTGAGCAGATGCTTGATTTGATTAAGATATATATAAAGGCT 817
    ||||| ..... |||
seq_name: gb_est12.AA788532

```

seq. documentation block: 263 bp mRNA EST 29-JUL-1998

LOCUS AA788532

DEFINITION alb02c9.r1 Aspergillus nidulans 24hr asexual developmental and vegetative cDNA lambda zap library Aspergillus nidulans cDNA clone alb02c9 5', mRNA sequence.

ACCESSION AA788532

VERSION AA788532.1 GI:2843185

KEYWORDS EST.

SOURCE Aspergillus nidulans.

ORGANISM Aspergillus nidulans

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiiales; Trichocomaceae; Emericella.

REFERENCE 1 (bases 1 to 263)

AUTHORS Kupfer,D., Gray,U., Hausner,J., Lal,H., Martin,W., Aramayo,R., Prade,R. and Roe,B.


```

104 ThrGlnLysLysLeuProGluArg.....LysTrpAlaIleValThrSe 118
      :::::||||| :::::||||| ||| ||| |||||:::
316 GTGAGAAATGATGATTGAAGATGATGAGCGAGCTGCGATTACCA 365
118 rGlySerProTyrLeuAlaPheSerTrpPheGlnThrIleLeuLysAsnV 135
      :::::||||| ||| ||| |||||:::
366 TGCTCCAGACCAATGCA.....GACCTCATTTCTCCAAAGC 403
135 aGlyLysProLysValPhe.....IleThrGlyPheAspValLys 148
      :::::||||| ||| ||| |||:::
404 TTGGCTCTCGAATTTCTTTTCATGCTGATTATTTGGTCATGATGATCA 453
149 AsnGlyLysProAspProGluGlyTyrSerArgAlaArgAspLeuLar 165
      :::::||||| ||||| ||| :::::
454 CATGCCAAGCCCTCATCCAGAACCCCTTGAAGAGCTTGAAGCTCTCA 503
165 gGlnAspLeuGlnLeuThrGlyLysGlnAspLeuLysTyrValValPheG 182
      : ::::: ||| :::::
504 ACCATCTAAGATCACACA.....TTTATTTTG 532
182 LuAspAlaProValGlyIleLysAlaGlyLysAlaMetGlyAlaIleThr 198
      |||||::: ||||| ||||| ||| |||
533 AGGATTCCTGTTCAAGAAATCAAGCTGCTGCGACGAGATGCTGTT 582
199 ValGlyIleThrSerSerTyrAspLysSerValIlePheAspAlaGlyAl 215
      :::::|||||::: :::::|||||::: |||
583 ATAGGTATATCTACTGAAACCCAGAAAGCTTGTATGAGAGCAAAACC 632
215 aAspTyrValValLysAsp 221
      ::::: |||
633 TGCCCTTTTGATTAAAGAT 651

```

seq_name: gb_gss30:A2682437

seq_documentation_block:

LOCUS A2682437 975 bp DNA GSS 14-DEC-2000
 DEFINITION ENTEN39TR Entamoeba histolytica Sheared DNA Entamoeba histolytica

genomic, DNA sequence.

ACCESSION A2682437 GI:11819583

VERSION A2682437

KEYWORDS GSS.

SOURCE Entamoeba histolytica.

ORGANISM Entamoeba histolytica.

REFERENCE 1 (bases 1 to 975)

AUTHORS Loftus,B., Van Aken,S. and Fraser,C.

TITLE Determination of clone end sequences from Entamoeba histolytica

COMMENT HMI:IMSS sheared DNA library

JOURNAL Unpublished (2000)

CONTACT: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: b.loftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared

DNA library

Seq primer: M13-Reverse

Class: shotgun

High quality sequence start: 81

Location/Qualifiers

1..975

/organism="Entamoeba histolytica"

/strain="HMI:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: pHOST1; Site_1: Bst I; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from Proth cultures of E. histolytica

using a method described by Clark and Diamond (Clark,

C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a light size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O., and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Borell, Oxford University Press, 1999).

BASE COUNT 321 a 134 c 125 g 395 t
 ORIGIN

alignment_scores:
 Quality: 126.50 Length: 233
 Ratio: 1.045 Gaps: 14
 Percent Similarity: 51.931 Percent Identity: 28.326

alignment_block:
 US-09-558-284-2 x A2682437/rev ..

Align seg 1/1 to reverse of: A2682437 from: 1 to: 975

```

10 LeuPheAspLeuAspGlyThrIleValSerThrThr.....ValAl 23
      ||||| ||||| ||||| ||||| ||||| |||||
667 TTATTGATTTAGACGCTACTTACTTACTGATACGAACCTTTATGACAC 618
23 aAlaGluLysAlaIleThrThrLysLeuLysCysTyrGluTyrGly.Valasp... 38
      : ::::: |||
617 AATTATCAAGAAATTTATTAATTTA.....TATGAGATGGAAGA 577
38 ..:::
576 ACTATGATTTGGGAACGAGAAACAAAGTTATGGGAAGTCAGCAGAAAT 527
39 .ProSerGluLeu...PheLysHisSerHisGlyAlaArgThrGlnGluV 54
      ||| ::::: ||||| ||||| |||||
526 GCCAATCCAAATATATCATTCACAAAC.....ATCATATTTCAAAAC 486
54 aLeuArgArgPhePheProLys.....LeuAspAspThr 65
      ::: ::::: |||
485 TTAAGAGAGAGTTTGTAAATAAACAATAAATAATATTAATAATAATA 436
66 AspAsnLysGlyValLeuAlaLeuGluLysAspIleAlaHisSerTyrLe 82
      ::: ::::: |||||
435 AACAAATTTAGATGTTAATTGAAGAAAGACATTTAGCACAACCTTG 386
82 uAspThrValSerLeuIleProGlyAlaGluLysLeuLeuSerLeuA 99
      ::: ||||| ||| ::::: |||
385 TGAAGAGAGTAAACCATTTCCAAACCATTTAGAAATTAATAACTTTT 337
99 sPValAspThrGluThrGlnLysLysLeuProGluLysTrpAlaIle 115
      ||||| ||||| |||
336 .....AACAAAAAGTTTA.....AAAGTGGCTATT 310
116 ValThrSerGlySerProTyrLeuAlaPheSerTrpPheGlnThrIle 132
      ::: ||||| ||||| |||||
309 GCTACAGTAGTGCACAAACAAAT.....TTGGAACCAAAAT 272
132 uLysAsn.....ValGlyLysProLysValPheIleThrGlyP 145
      : |||||
271 GAAAGAAGATCAAGAAATTAATCAATATAGTAGATGTTGGAG 222
145 heAsp.....ValLysAsnGlyLysProAspProGluGlyTyrSerArg 159
      ||| ::::: ||||| ||||| |||||
221 ATGATTTCTTCTGTTCACTCACTCAAAACCTGCACCTGATATTTATTA 172
160 AlaArgAspLeuLysArgGlnAspLeuGlnLeuThrGlyLysGlnAspLe 176
      ||| ::::: ||||| ||||| |||||
171 GCA.....GCCAAGCTTTGGAGAAACAAAGATAT 143
176 u...LysTyrValValPheGlnAspAlaProValGlyIleLysAlaGlyL 192
      : ||| ::::: ||||| ||||| |||||

```

```

142 GAGTAACCAATGTATTGACATGCATTAATGAGTAGAAGCAGCAT 93
192 ysal1amelcylalialcthrvalgilyl1elhhsereferlyrasplys 207
||||| |||||:::||:::||:::||:::||:::||:::||:::||:::||:::||
92 TGGCATTCCAGCGCATTTGACAATTCCTATACGACATCATATTTAAA 46

seq_name : gb_gss5.CNS06DY4

seq_documentation_block:
LOCUS      CNS06DY4          808 bp            DNA             GSS              05-APR-2001
DEFINITION T3 end of clone AR0AA017A03 of library AR0AA from strain CBS 732 01
ACCESSION  AL394418
VERSION    AL394418.1 GI:12145375
KEYWORDS   GSS: clone AR0AA017A03; RST AR0AA017A03TP1.
SOURCE     Zygosaccharomyces rouxii.
ORGANISM   Zygosaccharomyces rouxii.
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccaromycetales; Saccharomycetaceae; Zygosaccharomyces.
REFERENCE  1 (bases 1 to 808)
AUTHORS   de Montigny,J., Straub,M., Potier,S., Tekata,F., Dujon,B.,
            Wincker,P., Artiguenave,F. and Souciet,J.
TITLE     Genomic exploration of the hemiascomycetous yeasts : 8.
JOURNAL   Zygosaccharomyces rouxii
MEDLINE   FEBS Lett. 487 (1), 52-55 (2000)
REFERENCE  2 (bases 1 to 808)
AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Boitron-Pukharara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Leplingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozler-Kalogeropoulos,O., Potier,S.,
            Sarlin,M., Tekata,F., Toftano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
TITLE     Genomic exploration of the hemiascomycetous yeasts : 1. A set of
JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
REFERENCE  3 (bases 1 to 808)
AUTHORS   Direct Submission
TITLE     Submitted (28-Mar-2001) Genoscope - Centre National de Sequencage,
JOURNAL   2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
COMMENT   seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
            This BAC end sequence is part of a random genomic sequencing
            program of thirteen yeast species:
            . Saccharomyces bayanus var. uvarum,
            . Saccharomyces exiguus,
            . Saccharomyces servazii,
            . Zygosaccharomyces rouxii,
            . Saccharomyces kluyveri,
            . Kluyveromyces thermotolerans,
            . Kluyveromyces lactis var. lactis,
            . Kluyveromyces marxianus var. marxianus
            . Pichia angusta,
            . Debaryomyces hansenii var. hansenii,
            . Pichia sorbitophila,
            . Candida tropicalis and
            . Yarrowia lipolytica.
            Genomic inserts of 3 to 5 kb were prepared and both extremities
            were sequenced. See keywords for description of this sequence and
            for the sequence of the other extremity of this insert.
FEATURES
source     location/Qualifiers
            1..808
            /organism="Zygosaccharomyces rouxii"
            /strain="CBS 732"
            /db_xref="taxon:4956"
            /clone="AR0AA017A03"
            /clone_1id="AR0AA"
            /note="end : T3"
            complement(<3..>370)
            /note="similar to Saccharomyces cerevisiae ORF YIL053w [
            RHR2 ; DL-glycerol phosphatase ]
            2 putative frameshift(s)"

```

```

misc.feature                                /evidence=not_experimental
                                           /complement(<3..>358)
                                           /note="Similar to Saccharomyces cerevisiae ORF YER062c {
                                           /OR2 : DL-glycerol phosphate }
                                           /OR2 : DL-glycerol phosphate }
                                           /evidence=not_experimental
                                           /evidence=not_experimental

BASE COUNT      245 a      160 c      188 g      210 t      5 others
ORIGIN

alignment_scores:
      Quality: 123.00      Length: 119
      Ratio: 1.757      Gaps: 6
Percent Similarity: 58.824      Percent Identity: 33.613

alignment_block:
US-09-558-284-2 x CNS06DY4/rev ..

Align seg 1/1 to reverse of: CNS06DY4 from: 1 to: 808

9 CysLeuPheAspLeuPheAspGlyThrIleVal.SerThrThrValAlaIaIaG 25
|||||.....:|||||.....:|||||.....:|||||.....:|||||
316 TGCTCTTTGACGCTTGACCGTACCATCATCATCTCCCAANCCAGCTATGCG 267
25 IuIaSaIaTtPthrIuIaSaTyGtGluTyGly.....:.....: 36
|||||.....:|||||.....:|||||.....:|||||.....:|||||
266 CTGATCTTCGTAGAGA.....:.....:GAAATTCGCTAAGGACAAGCCTTAC 229
37 ValAspProSerClnLeuPheIuIaSaSerHisGlyAlaIaArgThrIleG 53
|||||.....:.....:|||||.....:|||||.....:|||||.....:
228 TTGCACTTCCTCCAAATACGTCAATCAACATCTCTACGGTTGGACCTTAACG 179
53 uValLeuArg.ArgPhePheProLyLeuAspAspPthrAspAsnLyGly 69
|||||.....:|||||.....:|||||.....:|||||.....:|||||
178 ACGCTTTCGCTAAGTTGCGTCCA.....:GACTACGCTGATGAAGAATAC 135
70 ValLeuAlaIeugIuIuIaSaPllIeAlaHisSerTyLeuAspPthrValSe 86
|||||.....:.....:|||||.....:|||||.....:|||||.....:
134 GCTACCAAGGTGACAGCTCAAAATCCAGACAATACGATCAAGTCCAT 85
86 rIeuIleProGlyAlaGluAsnLeuLeuSerIeuaSpValAspThrG 103
|||||.....:.....:|||||.....:|||||.....:|||||.....:
84 CGAAGTCGCCGGTGTACAGAGTTGTCTAAGCTTTGAACGCT.....: 42
103 IuThrIuIuIaSaLeuProGluArgLysTrpAlaIleValThrSerGly 119
|||||.....:.....:|||||.....:|||||.....:|||||.....:
41 .....:.....:TTGCCAAAGGAAATAATGGCGCTTGCTACTTCCGCT 6

120 Ser 120
|||
5 TTC 3

seq_name: gb_est102:BG589791

seq_documentation_block:
LOCUS BG589791 728 bp mRNA EST 12-APR-2001
DEFINITION ES1494633 P. infestans-challenged leaf Solanum tuberosum cDNA clone.
BPLI3C1 5' sequence, mRNA sequence.
ACCESSION BG589791
VERSION BG589791.1 GI:13607931
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
Asteridae; easterids I; Solanales; Solanaceae; Solanum.
1 (bases 1 to 728)
Zhang,P., Hernandez,M., Toruqvist,C.-E., Wirtz,U., Loukianov,A.,
Rangel,P., Haberlach,G.T., Cho,J., Chilingiro,A., Bougri,O., Bueli
,C.R., Renning,C.M., Helgeson,J. and Baker,B.
Generation of ESTs from Potato leaves Challenged with Phytophthora
infestans. Incomplete Reaction
Unpublished (2000)
JOURNAL

```

COMMENT

Contact: Cathy Renning
The Institute for Genomic Research
For clone info: Please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: M13F-R.
Location/Qualifiers

FEATURES

source

1..728

/organism="Solanum tuberosum"

/cultivar="Kennebec"

/db_xref="taxon:4113"

/clone="BPL1301"

/clone_lib="P. Infestans-challenged leaf"

/tissue_type="leaf"

/dev_stage="6 week old"

/lab_host="SOLR"

/note="Vector: pBluescript SK(-). Site_1: EcoRI; Site_2: XhoI; Whole plants were challenged with 450,000 sporangia/ml P. infestans US-1(US 940501) in Biotron (Madison, Wisconsin). Leaf tissue was collected at 1, 2, 5, 12, and 24 hours post-challenge and frozen in liquid nitrogen immediately upon removal. Kennebec plants showed no signs of HR. Kathadin plants (susceptible to P. infestans US-1) were used as controls and showed infection. NOTE: We cannot exclude the possibility that this sequence is actually derived from Phytophthora rather than potato."

BASE COUNT 190 a 148 c 177 g 213 t
ORIGIN

alignment_scores:

Quality: 121.00 Length: 203
Ratio: 1.210 Gaps: 7
Percent Similarity: 49.261 Percent Identity: 24.138

alignment_block:

US-09-558-284-2 x BG589791 ..

Align seg 1/1 to: BG589791 from: 1 to: 728

```

34 GltutrglyvalaspProsergluleuphellyshisserhisglialaar 50
   ::::::::::::::::::::
33 GATTATGGCGCTGCCACTGTCA..... 53
50 gThnglgluValleuargrarpherphepolysleuAspAspThrAspa 67
   ::::::::::::::::::::
54 .ACGATGCAATTCTTCGCAATTATACCAATGCTGCTGAC..... 95
67 snlysglyValleualaleuclulysaspilleaHisSertyrleuasp 83
96 .....
84 ThrValSerleuLeuProglialagiuanleu..... 95
105 AATATCAAGCTGTCCAGGGCCAAATCGATTGATTACCATTTGAGGG 154
96 .....LeuSerleuaspValasPthrgluThrglnlyslvsleuProg 110
155 TCATGGGTACCCATGGCATTTGGCGTCAATTTCTTAAGATCAACATAG 204
110 luarglystrPalalilevalThSerGlySerProtyrleuAlapheser 126
205 AGGCCCAA.....ATTTTCATCATGCAGCA 230
127 TrpPheglThrIleleuLysasnValglylvsProlvsvalPheleth 143
231 TCGAAGAAATCCCTTCACGCTATGTTGA..... 260
143 rglYpheaspValLysasnGlyLysProaspProgluglytyrSerargA 160
261 .CGACGCAAGTCAAGCTGAAAGCCATCTCCGGAATATTTCTTGAAG 309
160 laArgaspLeuuarGlnaspLeuGlnleuThrglyLysGlnaspLeu 176

```

```

310 CAGCTAAAGACTCAACATGAT.....CATCC 338
177 LysTrValValIpegluaspAlaProValGlyIleLysAlaGlyLysAl 193
   ::::::::::::::::::::
339 AGCTTCTTCATTTGAGATTCATACACAGGTGTACTCTGCTAAGGC 388
193 aMetGlyAlaIleThryValGlyIleThrserserTyraSpLysServall 210
   ::::::::::::::::::::
389 TGTGTAATGACAGTACTGTGTCATCATCTTCGCAACACATGATC 438
210 eupheaspAlaGlyAlaAspTyValValGlyaspLeuThrglnValser 226
439 TTTAT...ACTTGTGATGATGATCAATTCCTCTAGATTGGCAA 485
227 ValValLys 229
486 CTGCAAAAG 494

```

seq_name: gb_gss30:AZ682029

seq_documentation_block:

LOCUS AZ682029 852 bp DNA GSS 14-DEC-2000
DEFINITION ENTROP38TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
ACCESSION AZ682029 genomic DNA sequence.
VERSION AZ682029.1 GI:11819175
KEYWORDS GSS.

SOURCE

Entamoeba histolytica.

Eukaryota; Entamoebidae; Entamoeba.

1 (bases 1 to 852)

Loftus/B., Van Aken, S. and Fraser, C.
Determination of clone end sequences from Entamoeba histolyticaHMI:IMSS sheared DNA library
Unpublished (2000)

JOURNAL

COMMENT

Contact: Brendan J Loftus
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 15
High quality sequence stop: 836.
Location/Qualifiers

FEATURES

source

1..852

/organism="Entamoeba histolytica"

/strain="HMI:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: pHOSt; Site_1: Bat I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 287 a 120 c 114 g 331 t
ORIGIN

alignment_scores:

Quality: 121.00 Length: 207
Ratio: 1.198 Gaps: 9
Percent Similarity: 48.792 Percent Identity: 24.638

alignment block:
US-09-558-284-2 x A2682029/rev ..

Align seg 1/1 to reverse of: A2682029 from: 1 to: 852

```

28 TrpThrIleuSerCysTyrGluTyrGlyValAspProSerGluLeuPheIly 44
   ||||| ..... |||||
851 TATTCACATTTGTTACTCACCAATAT.....TTAGATGA 820
   ||||| ..... |||||
44 ShisSerHisGlyAlaArg...ThrGlnGluValLeuArgArgPhePhe 60
   ||||| ..... |||||
819 ATATGCTTAATGAGCTAAATTACTTATGAAATTAATAAACAAATGATGG 770
   ||||| ..... |||||
60 rOlYsIleuAspAspThrAspAsnLysGlyValLeu..... 71
   ||||| ..... |||||
769 GAAACATATATTGAGCTTGCTACTAAATGCTTAATGATATTTCATATTT 720
   ||||| ..... |||||
72 .....AlaLeuGluLysAspIleAlaHis.....Se 80
   ||||| ..... |||||
719 AATGACACACTTGACATGCAATACAAATATAAATTCATCATTAATTA 670
   ||||| ..... |||||
80 rTyIleuAspThrValSerIleuIleProGlyAlaGluAsnLeu..... 95
   ||||| ..... |||||
669 ATTATGGTCAACAGCTAAACCAATACAGAGCAATGACGATTTTAATTT 620
   ||||| ..... |||||
96 .....LeuSerIleuAspValAspThrGluThrGlnLys 106
   ||||| ..... |||||
619 ATTTTAAAAAATCATCATTCCTATGCTCTTGCTACTTCAACCACTAAA 570
   ||||| ..... |||||
107 LysLeuProGluArgLysTrp..... 113
   ||||| ..... |||||
569 TCTGCTCTTTGACAAAATAATGCTTAAACCAAGAGATGCTCAATTTATTT 520
   ||||| ..... |||||
519 TGATGCAATTTGCTCTTGAGATGATCTCAT..... 489
   ||||| ..... |||||
129 LuthrIleLeuLysAsnValGlyLysProLysValPheIleThrGlyPhe 145
   ||||| ..... |||||
489 ..... 489
146 AspValLysAsnGlyLysProAspProGluGlyTyrSerArgAlaArgAs 162
   ||||| ..... |||||
488 .....GTAAGAAGAGCTAAACCAATTCACAAATATTTTACATGCTGCTCA 442
   ||||| ..... |||||
162 pLeuLeuArgGlnAspLeuGlnLeuThrGlyLysGlnAspLeuLysTyr 178
   ||||| ..... |||||
441 TTTACTT.....GGATGCTACTGACATGCAACAG 413
   ||||| ..... |||||
179 ..ValValPheGluAspAlaProValGlyIleLysAlaGlyLysAlaMet 194
   ||||| ..... |||||
412 CAATTTGATTTGAAAGATGCACTGCTGCTTCAAGCAGCAATTCATCT 363
   ||||| ..... |||||
195 GLYAlaIleThrValGlyIle 201
   ||||| ..... |||||
362 GAGAGCTTATCTGATGCTATTT 342
   ||||| ..... |||||

seq_name: gb_est30:AU093031
seq_documentation_block:
LOCUS AU093031 635 bp mRNA EST 14-JUN-2000
DEFINITION AU093031 Rice callus Oryza sativa subsp. japonica cDNA clone C61715
, mRNA sequence.
ACCESSION AU093031
VERSION AU093031.1 GI:8528216
KEYWORDS EST
SOURCE Oryza sativa subsp. japonica.
ORGANISM Oryza sativa subsp. japonica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Erihartoideae; Oryzeae; Oryza.
REFERENCE
1 (bases 1 to 635)
AUTHORS
Sasaki, T. and Yamamoto, K.
TITLE
Rice cDNA from callus (2000)
JOURNAL
Unpublished (2000)
CONTACT
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT "RGP".
C61715_82.

FEATURES
source
Location/Qualifiers
1..635
/organism="Oryza sativa subsp. japonica"
/strain="cultivar Nipponbare, sub-species japonica"
/db_xref="taxon:39947"
/clone="C61715"
/clone_lib="Rice callus"
/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site
of pBluescript II SK+ phagemid.
5 others

BASE COUNT 182 a 132 c 144 g 172 t
ORIGIN

alignment_scores:
Quality: 117.50 Length: 82
Ratio: 2.350 Gaps: 1
Percent Similarity: 60.976 Percent Identity: 35.366

alignment_block:
US-09-558-284-2 x AU093031 ..

Align seg 1/1 to: AU093031 from: 1 to: 635

```

140 ValPheIleThrGlyPheAspValLysAsnGlyLysProAspProGlu 156
   ||||| ..... |||||
51 CTGCTCGTAATCGGAGGAGCTGACCGAGCAACCATTCCTCGATGCC 100
   ||||| ..... |||||
156 yTySerArgAlaArgAspLeuLeuArgGlnAspLeuGlnLeuThrGly 173
   ||||| ..... |||||
101 TTACCTCAAGGCT.....CTTGAGCTTACTGCTG 129
   ||||| ..... |||||
173 ysgInAspLeuLysTyrValValPheGluAspAlaProValGlyIleLys 189
   ||||| ..... |||||
130 CATCACTGATCACACATTCATTTTGGAGCTCCGGCTCAGGAGATTCGA 179
   ||||| ..... |||||
190 ALaGlyLysAlaMetGlyAlaIleThrValGlyIlePheSerSerTyrAs 206
   ||||| ..... |||||
180 GCCGCTGAGCTGCAGCAATGTTCTGCTAGTTGCTGACCACTAGAGATCC 229
   ||||| ..... |||||
206 pLysSerValLeuPheAspAlaGlyAlaAspTyrValValLysAsp 221
   ||||| ..... |||||
230 AGAGAAAGTTCTGCAGATGCGAGGACGACGATCTGTATTAAGAT 275
   ||||| ..... |||||

seq_name: gb_est50:AW696074
seq_documentation_block:
LOCUS AW696074 655 bp mRNA EST 15-JUN-2000
DEFINITION NE102C08ST1F1065 Developing stem Medicago truncatula cDNA clone
NE102C08ST 5', mRNA sequence.
ACCESSION AW696074
VERSION AW696074.1 GI:7570836
KEYWORDS EST
SOURCE barrel medic.
ORGANISM Medicago truncatula

```

REFERENCE
AUTHORS

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
1 (bases 1 to 653)
He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell
C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,G.D. and Dixon
R.A.

TITLE
JOURNAL
COMMENT
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula stem library
Unpublished (2000)
Contact: Dixon RA
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7302
Fax: 580 221 7380
Email: radixon@noble.org
Insert Length: 653 Std Error: 0.00
Plate: 102 row: C column: 08
Seq primer: TCACACAGGAACAGCTATGAC.
Location/Qualifiers

FEATURES
source
1. 655
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF102C08ST"
/clone_lib="Developing stem"
/tissue_type="stem"
/dev_stage="Pooled developmental"
/note="Vector: Lambda Zap; Contains a mixture of
internodal stem segments"

BASE COUNT 188 a 127 c 149 g 191 t
ORIGIN

alignment_scores:
Quality: 117.00 Length: 205
Ratio: 1.093 Gaps: 9
Percent Similarity: 52.195 Percent Identity: 26.829

alignment_block:
US-09-558-284-2 x AM696074 ..

Align seg 1/1 to: AM696074 from: 1 to: 655

```

10 LeupheaspLeuaspGlyThrIleValSerThrThrValAlaIaIaIu 26
   |||||.....|.....|.....|.....|.....|.....|
96 TTGTTGATGATAGACGAACTTTTGATTCTGACCAATCCATCTTA 145
   |||||.....|.....|.....|.....|.....|.....|
26 sAlaTPrThLysLeuCysTyrGluTyrGlyValAsp..... 38
   |||||.....|.....|.....|.....|.....|.....|
146 TGCTTTTCGTGAATGCTTCAGAGATTGCTTCAATGAGGCAATCTTA 195
   |||||.....|.....|.....|.....|.....|.....|
39 ..ProserGluLeuPheLysHisSer...HisGlyAlaIaIaGthrGlu 53
   |||||.....|.....|.....|.....|.....|.....|
196 TACACAGAAATTTTATTGCGACTGTTGCTGCAGCAACCAATGATGAC 245
   |||||.....|.....|.....|.....|.....|.....|
54 ValLeuAlaIaIaGpPhePheProLysLeuAspPThrAspLysGly 70
   |||||.....|.....|.....|.....|.....|.....|
246 ATGCTTTGATCTCTTCCC.....GGATCGGAGAGAGCTTTAA 289
   |||||.....|.....|.....|.....|.....|.....|
70 lleuAlaIaIaGluLysAspIleAlaHisSerTyrLeuAspPThrValSer 87
   |||||.....|.....|.....|.....|.....|.....|
290 GTTGTAGAGATAGGAAGCCATGTTGCGAGATTACGACGACCAAC 339
   |||||.....|.....|.....|.....|.....|.....|
87 euIleProGluAlaGluAsnLeuLeuSerLeuAspValAspThrGlu 103
   |||||.....|.....|.....|.....|.....|.....|
340 TGAAGCC.....TTGAATGCTCTGAT.....AAA 365
   |||||.....|.....|.....|.....|.....|.....|
104 ThrGluLysLysLeuProGluArg.....LysTrpAlaIleValThr 118
   |||||.....|.....|.....|.....|.....|.....|
366 GTGAGGAATGATGATGACATCGTGAAGCCAGCTGCAATTCACA 415
   |||||.....|.....|.....|.....|.....|.....|
118 rGlySerProTyrLeuAlaIaPheSerTyrPheGluThrIleLeuLysAsn 365

```

```

416 TGCTCCAGAACCAATCA.....GAACCTATTCTCCAAAGC 453
   |||||.....|.....|.....|.....|.....|.....|
135 alGlyLysProLysValPhe.....IleThrGlyPheAspValLys 148
   |||||.....|.....|.....|.....|.....|.....|
454 TTGCTCTCAGATTCTTTCATCCTCTTATTATGTGATGATGAA 503
   |||||.....|.....|.....|.....|.....|.....|
149 AsnGlyLysProAspProGluGlyTyrSerArgAlaArgAspLeuArg 165
   |||||.....|.....|.....|.....|.....|.....|
504 CATGCCAAGCCTCATCCAGAACCTACTTGAAAGGCTTGAAAGCTGCAA 553
   |||||.....|.....|.....|.....|.....|.....|
165 gGluAspLeuGlnLeuThrGlyLysGluAspLeuLysTyrValValPheG 182
   |||||.....|.....|.....|.....|.....|.....|
554 AGCATCTAAGATCACACA.....TTTATTATTG 582
   |||||.....|.....|.....|.....|.....|.....|
182 LuAspAlaProValGlyIleLysAlaGlyLysAlaMetGlyAlaIleThr 198
   |||||.....|.....|.....|.....|.....|.....|
583 AGGATTTCTGTTTCAGGAATCAAGCTGCTGCTGCGAGGATGCCGTG 632
   |||||.....|.....|.....|.....|.....|.....|
199 ValGlyIleThrSer 203
   |||||.....|.....|.....|.....|.....|.....|
633 ATAGTATATCTACT 647
   |||||.....|.....|.....|.....|.....|.....|

```

seq_name: gb_est30:AU100758

seq_documentation_block:

LOCUS AU100758 715 bp mRNA EST 22-AUG-2000
DEFINITION AU100758 Rice callus Oryza sativa subsp. japonica cDNA clone C12953
ACCESSION AU100758
VERSION AU100758.1 GI:9865008
KEYWORDS EST

SOURCE
ORGANISM
Oryza sativa subsp. japonica.
Oryza sativa subsp. japonica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE
1 (bases 1 to 715)
Sasaki,T. and Yamamoto,K.
Rice cDNA from callus (2000)

TITLE
JOURNAL
COMMENT
Unpublished (2000)
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT = "RGP".
C12953-82.

FEATURES
source
1. 715
Location/Qualifiers

/organism="Oryza sativa subsp. japonica"
/strain="cultivar Nipponbare, sub-species japonica"
/db_xref="taxon:39947"
/clone="C12953"
/clone_lib="Rice callus"
/note="Vector: pBluescript II SK+; site_1: SalI; site_2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site
of pBluescript II SK+ phagemid."
BASE COUNT 209 a 152 c 158 g 194 t 2 others
ORIGIN

alignment_scores:
Quality: 115.50 Length: 90
Ratio: 2.100 Gaps: 1
Percent Similarity: 61.111 Percent Identity: 32.222
alignment_block:

THIS PAGE BLANK (USPTO)